PILOT SIGNAL DESIGN FOR ESTIMATION OF SPARSE CHANNELS WITH APPLICATION TO COOPERATIVE SYSTEMS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Matthew Daniel Sharp

January 2011

PILOT SIGNAL DESIGN FOR ESTIMATION OF SPARSE CHANNELS WITH APPLICATION TO COOPERATIVE SYSTEMS

Matthew Daniel Sharp, Ph.D.

Cornell University 2011

In recent years, there has been a resurgence in the interest of using HF radio band (3-30 MHz) for military, government, and emergency applications. In order to allow for more reliable mobile HF radio networks, the use of distributed cooperative communication schemes is appealing. In fact, the large HF wavelength prevents one from obtaining diversity from antennae located on a single portable radio. Hence cooperative schemes can potentially bring the benefits of MIMO systems to portable HF radios, allowing for diversity gains as well as potential reductions in transmit power.

Our work is focused on the receiver design for cooperative HF radios. The motivation for our work is as follows. Using distributed cooperative communication schemes in OTH-HF propagation environments introduces several additional complications in comparison with traditional MIMO systems. Different from a traditional MIMO system, there is additional overhead required to organize and synchronize the cooperative radios' transmissions. Our work focuses on the distributed cooperative communication scheme introduced in [82] which reduces the overhead by introducing randomization into the coding scheme, but does not solve the synchronization problem. In terms of efficiency, it is best to allow cooperative nodes to transmit concurrently, using space-time coding techniques, as done in the randomized protocol proposed in [82]. But several authors argue against such solutions, given the challenge of synchronizing the different cooperative radios

participating in the transmission.

Our work investigates the effects of imperfect synchronization among cooperative nodes, resulting in dispersive effects that can be captured by an equivalent MIMO channel. The key contribution is in proposing the use of compressed sensing (or sparse signal recovery) techniques to deal with these channels at the receiver. We first consider designs that are sufficiently informative for MIMO systems that can be appropriately modeled as sparse. After introducing noise and modeling error, we examine the performance of an ideal sparse channel estimation method, leading to a metric we call *localized coherence*, and modified training designs. These results are then applied to the asynchronous distributed cooperative communication scheme.

BIOGRAPHICAL SKETCH

Matthew Sharp was born January 27th, 1981. He received his B.S. Degree in Electrical Engineering in 2003 from the George Washington University, Washington, DC, and his M.S. in Electrical Engineering in 2005 from Bucknell University, Lewisburg, PA. He joined Cornell University in 2005 as a graduate student, where he obtained his Ph.D. in 2010. His research interests include signal processing and wireless communications.

This document is dedicated to Claudia, David, Marshall, Becca and Dee Dee.

ACKNOWLEDGEMENTS

It has been a pleasure to work in a number of exciting and intellectually stimulating environments, with a wonderful array of mentors and colleagues. I would like to thank Professor Richard Kozick for stimulating my interesting in signal processing. He introduced me time the world of research, and encouraged me to further pursue my studies in a graduate program.

I thank my Advisor, Professor Anna Scaglione, for her support throughout my studies. Over the years she has imparted to me part of her knowledge. Her way of thinking and her guidance have had a significant impact on my approach to scientific research.

I am immensely grateful to all people who supported me during these years, including my family, friends, and relatives. It is because of them that my years spent on this work have been such a delight. I would also like to thank the research team at Telcordia Technologies for their feedback on this research, as well as providing a level of practicality to this research. Finally, I would like to thank Cornell University and the Office of Naval Research for supporting me during my research.

Part of this thesis represents joint work with my Advisor, Professor Anna Scaglione (UC Davis) and Professor C. R. Johnson, Jr (Cornell).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Since the early 1920's, radio communication services have been using high-frequency (HF) skywave propagation in order to establish long-range communication links. HF propagation is usually distinguished between ground wave and skywave propagation. Ground wave propagation occurs when the receiver is sufficiently close to the transmitter such that it is able to receive a portion of the signal traveling along the ground. The range for ground wave communication depends on the terrain between transmitter and receiver, the antenna used at the transmitter, as well as many other factors, but is typically on the order of a few tens of miles. On the other hand skywave propagation occurs when the transmitting signal radiates upwards and reflects off various layers of the ionosphere. Due to the reflections generated in skywave propagation, long distance, including over-the-horizon (OTH), communication is possible and can yield ranges of up to 1800 miles for a single hop to 3700 miles for two hops. It is thus no surprise that HF propagation was the primary method of long-distance communication for military forces during World War II, as a result of its ability to link air, land, and sea forces.

With the introduction of satellite communication in the 1960's, there was a decline in interest in HF radio. Satellites allowed for larger bandwidths and higher transmission speeds, and HF radio was used primarily as a backup. However, over time, it became clear that satellites had limitations. The military was concerned with jamming and possible physical damage to the satellites, the expense of the supporting infrastructure, and the maintenance of these satellite systems. This led to a resurgence of interest for HF radio for military applications as well as government and emergency applications.

When using low-powered man-portable radios, ground wave propagation is typically limited to 10-20 miles and in order to allow for OTH-HF radio communication, more powerful radios are needed. In these situations, cooperative communication may be desirable, where multiple man-portable radios can cooperatively transmit a common message to a single OTH destination, introducing diversity and robustness into the system.

Cooperative communication schemes are based on the principles of multiple-input multiple-output digital communications systems. The advantages of MIMO systems are widely acknowledged, and include significant increases in data throughput and link range without additional power or bandwidth, [35, 93]. One of the commonly used schemes in MIMO systems is that of the diversity scheme. Diversity schemes attempt to improve reliability of a message signal by using multiple communication channels with different characteristics. The assumption is that different channels experience different levels of interference and fading. Redundant information regarding the message is transmitted on each of the different channels and combined at the receiver, combatting the fading and making the system more robust.

A common diversity scheme is that of Space-Time Coding (STC) [55, 85, 86], but its use for the HF channel is problematic since the separation between antennae required to harvest diversity gains is on the order of several hundred meters and cannot be achieved through a single man-portable radio. However, cooperative schemes can be used as an effective means to obtain the gains of MIMO systems for the OTH-HF channel, by employing multiple low-powered man-portable radios in a collaborative fashion each functioning as one of the antenna elements of the MIMO system. One of the main goals in employing a cooperative system is to

minimize the overhead necessary for the nodes to operate in a collaborative fashion. In [82], a randomized cooperative communication scheme was introduced, requiring the exchange of little or no overhead between the cooperating radios. The scheme exploited multiple sufficiently spaced radios by introducing randomization into the space-time coding operation. As such, this scheme was also adaptive with respect to the number of radios cooperating.

In order to harvest the diversity gains from this cooperative communication scheme, accurate knowledge of the equivalent channel at the receiver is necessary. Typically, the channel is estimated at the receiver through the use of known training or pilot sequences inserted into the message sequence. The channel estimation is complicated in OTH-HF communications as a result of the reflections of the signal off the various layers of the ionosphere. These OTH-HF channels are typically characterized by significant Delay and Doppler spread [97], depending on the time of day a latitude [1, 3]. In applying the randomized scheme discussed in [82] to OTH-HF communication, the potential asynchrony among cooperative radios further complicates the channel estimation process by introducing larger delay and time-variations into the channel.

In the cooperative literature, authors often cite the synchronization issues as being possibly so severe that cooperative radios should not be allowed to transmit concurrently. Therefore, they advocate selecting a single relay at a time, for example [14]. Unfortunately, these schemes lead to other problems, such as increased overhead necessary to select a relay, and reduced diversity and coding gains. Motivated by the interest in applying the randomized space-time codes proposed in [82] for tactical OTH-HF communications, our research has been focused on solving the problems introduced by these potential effects at the receiver end, through the

study of more powerful channel estimation methods that exploit the structure of the cooperative channel.

The problem of estimating the sampled channel impulse-response of a time-varying channel involving HF radio links has been widely studied and motivated many of the channel estimation schemes used in today's wireless communications [25, 43, 47]. Channel estimation/equalization schemes can typically be characterized as either batch or adaptive. Batch channel estimation schemes operate on a block by block basis, and often assume block-fading, i.e. the impulse varies only from block to block. On the other hand, adaptive channel estimation schemes typically operate on a symbol-to-symbol basis, constantly refining their estimate of the channel, and as such can track slow time-variations in the channel. The problem arising in these distributed cooperative communication schemes on HF channels, is that due to asynchrony among the cooperative radios and the narrow bandwidth of the HF channel, multiple carrier offsets can cause time-variations in the frequency-selective channel that violate the block-fading assumption of the batch channel estimation methods, and vary too fast for accurate tracking and prediction by simple adaptive methods. These channels call for more advanced channel estimation techniques.

The channel between each cooperative radio and the receiver is a result of the physical propagation paths as well as a delay and carrier offset due to the asynchrony between the radio and the receiver. In HF skywave propagation, the number of propagation paths is a function of the number of reflecting layers of the ionosphere, and according to [3] is well approximated by two paths. Each of these paths can be approximately characterized by a particular delay and Doppler shift including the effects of asynchrony. Assuming the cooperative radios are sufficiently

spaced such that there channels are independent, the total number of propagation paths contributing to the equivalent randomized channel at the receiver, when using [82], is on the order of $2T$, where T is the number of cooperative radios while the total number of parameters in the discrete-time channel impulse response is on the order of KL where K is the number of samples the channel is observed over and L is the memory of the discrete time channel. When $2T \ll KL$, we say that the channel can be modeled as sparse, and in this work we attempt to exploit this sparsity to estimate the channel.

In the past several years there has been much research in the identification and recovery of sparse signals from a limited number of observations [17, 19, 21, 28, 29, 37, 41, 90, 91]. The main idea behind this research is that the additional knowledge that the signal to be recovered is sparse over an appropriately defined basis, allows, in certain situations, the identification of this signal in underdetermined linear systems. The application of these methods to the problem of channel estimation allows accurate estimation of the doubly-selective channels that can be sufficiently modeled as sparse over a suitably defined overcomplete basis for the channel space. The question remains as to the amount and type of training necessary to identify these channels, and further, how to design training sequences that are robust to noise and the inherent modeling error present in these applications.

The main contribution of this work is that of generalizing the basic theory of sufficiently informative inputs [61], to determine a class of training signals that ensures system identifiability. WThe insights gained from these sufficiently informative inputs were used to find robust training designs.

1.1 Thesis Statement

In this thesis we study the design of sufficiently informative inputs, namely those that allow identifiability [61], for systems that can be modeled as sparse over appropriately designed bases. We propose several sufficiently informative input designs for MIMO systems that can be modeled as sparse. These designs are further modified to be more robust in the presence of noise and modeling error, and a metric, called localized coherence, is derived to allow comparison of these designs in terms of their performance.

We apply these sparse signal recovery techniques to the channel estimation in the asynchronous randomized cooperative communication scheme, and propose training designs based for such systems. We exploit the structure of the cooperative communication system, proposed modified sparse signal recovery techniques to further improve the performance of these schemes. We finally show that these training designs and channel estimation techniques allow us to achieve performance gains in comparison to traditional channel estimation techniques as well as non-cooperative systems.

1.2 Contribution

The main results we obtained are summarized below.

- **Sufficiently Informative Inputs:** We demonstrate that for the identification of SISO time-selective and/or frequency-selective systems, a certain amount of temporal and/or spectral support in the input (or probing) signal allows for identification of the system response. For MISO systems, we

demonstrate that a certain amount of spatial, temporal and spectral support in the probing signal allows for identification of the system response, while for MIMO systems, exploiting the multiple snapshots at the receiver array allows for a great reduction in the amount of resources used in the probing signal for identification of the system response.

- **Localized Coherence:** Based on an ideal sparse recovery algorithm in the presence of noise, we derive a metric called localized coherence, that can be used to compare probing signal designs in terms of their channel estimation performance. Based on this metric, we obtain modified designs of sufficiently informative inputs that allow for improved performance in the estimation of the system parameters in the presence of noise and modeling error.

- **Randomized Cooperative Communication:** We derive the obtainable diversity gain for randomized cooperative systems in frequency-selective channels, demonstrating that diversity can also be harvested from the frequency-selectivity of the channel. We then discuss and compare several different coding schemes for this cooperative communication scheme, as well as demonstrate that diversity gains can still be obtained when using this cooperative communication scheme in OTH-HF propagation environments. This is followed by applying the sufficiently informative input and localized coherence results to the asynchronous OTH-HF cooperative communication scheme, to design training designs for such systems. We demonstrate that diversity gains are still achievable in these systems when using sparse channel estimation techniques.

1.3 Outline

The thesis is organized as follows: Chapter 2 discusses the conditions for sufficiently informative inputs for MIMO systems that can be modeled as sparse, and proposed several sufficiently informative probing signal designs for these systems. In Chapter 3 we consider the addition of noise and modeling error to the system, and examine the performance of an ideal sparse channel estimation method in order to derive the metric of localized coherence. Based on this metric, modifications are made to the probing sequence designs in Chapter 2, resulting in more robust designs. In Chapter 4 we discuss the randomized cooperative communication scheme, deriving the obtainable diversity in frequency-selective channels, followed by the extension of the channel model to OTH-HF channels as well as asynchronous radios. We also compare the performance of several coding schemes through simulation. In Chapter 5 we extend the model to include non-negligible carrier offsets and apply the results of Chapter's 2 and 3 to design training sequences for these channels.

CHAPTER 2

SUFFICIENTLY INFORMATIVE EXCITATION FOR SPARSE LINEAR ESTIMATION PROBLEMS

2.1 Motivation and Related Work

A well-established concept in classical input-output system identification literature is that of *sufficiently informative input data*, allowing discrimination between all models in a particular model set. Consider the model set of linear systems with a rational transfer function $G(q) = H(q)/A(q)$. Assuming $H(q)$ and $A(q)$ are polynomials with degree at most $N_h - 1$ and N_a, we restate the following theorem from [61]:

Theorem 2.1.1 (Chapter 13, [61])**.** *Consider a set of single-input single-output (SISO) models given by:*

$$G(q) = \frac{\sum_{l=0}^{N_h-1} h[l]q^{-l}}{1 + \sum_{m=1}^{N_a} a[l]q^{-l}}$$

Then an open-loop experiment with an input that is persistently exciting of order $N_h + N_a$ is sufficiently informative (SI) with respect to the model set.

The difference $\triangle G(q) = G_1(q) - G_2(q)$ between any two models in the set consists of a numerator polynomial[1] with at most $N_h + N_a - 1$ zeros on the unit circle. In essence, the theorem states the following: an input signal whose spectrum contains at least $N_h + N_a$ nonzero points on the unit circle cannot be filtered to zero by $\triangle G(q)$; hence, no two models $G_1(q)$ and $G_2(q)$ can ever produce the same

[1]A simple computation shows that the numerator of $\triangle G(q)$ is expressed as $H_1(q)A_2(q) - H_2(q)A_1(q)$.

output if the input signal's power spectrum has at least $N_h + N_a$ nonzero values, i.e. it is *persistently exciting of order* $N_h + N_a$, [61].

We now restrict the model set and assume that the polynomials $H(q)$ and $A(q)$, with degrees such that $N_h + N_a - 1 < P$, are sparse in the sense that they have at most $S_h \ll N_h$ and $S_a \ll N_a$ nonzero coefficients, respectively. It follows that the numerator polynomial of $\triangle G(q)$ can have at most $\min(2S_hS_a, N_h + N_a)$ nonzero coefficients. Using Chebotarev's theorem, [83, 84], it can be shown that a sparse polynomial with at most S nonzero coefficients and degree less than P, with prime P, can have at most $S - 1$ zeros in the set of the P-th roots of unity. This result is the foundation for much of the current sparse signal recovery literature, [20, 21], and our objective is to revisit it in the context of *system identification*. In fact, applying this to the concept of SI inputs, if $S_hS_a < (N_h + N_a)/2$, this polynomial can have at most $2S_hS_a - 1$ zeros on the set of the P-th roots of unity. Thus an input signal that has at least $2S_hS_a$ nonzero points on the P-th roots of unity is SI for these sparse models. In summary, we have the following:

Lemma 2.1.2. *For a model $G(q)$ with at most $S_h + S_a$ non-zero unknown parameters out of $N_h + N_a$, such that $S_hS_a < (N_h + N_a)/2$, a sufficiently informative input need only have a spectrum consisting of at least $2S_hS_a$ nonzero points on the P-th roots of unity, where $P \geq N_h + N_a$ and prime.*

Thus, the additional knowledge that the signal is sparse, allows the number of nonzero spectral components to be reduced from $N_h + N_a$ to $2S_hS_a$; however these spectral lines are now restricted to the set of frequencies $\{0, 1/P, ..., (P-1)/P\}$. If $2S_hS_a \geq N_h + N_a$, the system is no longer sparse and again $N_h + N_a$ nonzero spectral points will be SI.

Our aim is to extend this result to a wider class of linear systems that arise due

to multi-path propagation. As done in most receivers, to reduce the complexity, we assume that the estimation and detection of symbols is decoupled. Our objective is to find a class of sufficiently informative pilot signals for sparse models that closely approximate a real time-varying multi-input multi-output frequency-selective channel. The challenging cases we are interested in are those of doubly-selective channels where due to the effects of Doppler and carrier offsets, the system is no longer accurately modeled as stationary impulse responses. The setup of Lemma 2.1.2 is insufficient to capture the models and a refinement of Lemma 2.1.2 will prove necessary even in the case of multi-path.

Optimal pilot design and placement for frequency-selective channels is considered in [4, 27, 68, 95]. Doubly selective channels have been known to be more challenging since the early studies by Bello [11]. A popular approach consists of modeling each of the time-varying taps of the channel as an autoregressive (AR) process, tracking the channels through the use of a Kalman filter [48, 92]. An alternative deterministic approach uses instead a basis expansion model (BEM), e.g. [16, 39], where each channel impulse response tap is represented as a linear combination of a finite set of time-varying basis functions whose coefficients can be retrieved with common linear parameter estimation techniques [10, 49, 62]. In [49], the authors consider the design of pilots that minimize a lower bound on the MSE of the linear MMSE channel estimate within the framework of affine-precoding. A combination of BEM and the stochastic approach has been proposed in [44]. One drawback of BEM is that the length of observations needed to estimate the channel is multiplicative in the number of channel taps and basis functions, [62].

In many situations [50, 57, 65, 76] (underwater acoustic channels, HDTV, and residential UWB channels, Radar) the system impulse response can be described

with less than half of the parameters through a linear combination of an appropriate set of basis functions (i.e. it is *sparse*). The choice of basis differs from BEM in that it is potentially a large overcomplete basis; in fact the emphasis is not placed on reducing the dimension of the space spanned by the basis but rather the number of coefficients required to represent the system. This has motivated a class of approaches for frequency-selective block fading channels that have recently been proposed in [7, 26, 36, 79], for doubly-selective SISO channels in [9, 45, 57, 60, 87], and for the similar Angle of Arrival (AOA) estimation problem in [63]. In [36], the problem was similar to the BEM approach, but one where the channel impulse response is modeled as the linear combination of delayed signals, whereas in [7, 26] the impulse response was itself a sparse vector.

By reversing the roles of the system and signals one can also find great similarity between our problem and that posed in [15, 94] pertaining to the reconstruction of signals with finite rate of innovation. In fact, an example of a signal with finite rate of innovation is a weighted sum of Dirac pulses, which is analogous to our multi-path model. Our results apply to the *sampling kernel* in [94].

The work in [8, 9, 45, 87] more closely relates to our results on doubly-selective channels. In [45] the Alltop sequence is proposed as a radar pulse, and it is argued that the identification of targets backscattering the pulse in time and frequency can be mapped into a Compresses Sensing (CS) problem satisfying a certain restricted isometry property (RIP). In [8,9], the identification of sparse doubly-selective (and MIMO doubly-selective) channels is examined. In these works the authors are focused on satisfying sufficient conditions for identifiability that are based on mutual coherence [28] and the so called RIP [22]. The linear expansion of the system is obtained by uniformly quantizing the delay-Doppler domain (or delay-Doppler-angle

domain) at a rate that is the reciprocal of the bandwidth and the signal duration (the number of array elements), respectively. The RIP guarantees the robust recovery of channels that are sparse in this sampled domain. In [87], modeling errors are handled by augmenting the support of the representation. In [88], the effects of modeling errors are further combated by optimizing the basis expansion with respect to the Doppler parameters, increasing the block length used for estimation, and in [32] the block sparsity structure of the channel is used. A similar extension to MIMO multi-carrier systems is proposed in [33].

2.1.1 Contribution and Motivation

The notion of *sufficiently informative* (SI) inputs [61], or persistently exciting signals, has been widely employed in the system identification literature to formally classify signals that guarantee uniqueness of the output, for a class of systems considered. The novel aspect of our work is in finding a class of SI inputs for an important class of linear systems whose impulse response description can be shrunk to less than half the number of parameters by resorting to a basis expansion. We refer to this class of systems as *sparse systems* (c.f. Sec. 2.2).

As previously noted, deterministic designs or random inputs that ensure identifiability with high probability proposed by others, see e.g. [9, 45, 87], are focused on satisfying a restricted isometry property (RIP) [20] for a fixed resolution on the parameter space directly tied to the number of orthogonal dimensions of the signal. The designs of SI inputs we discuss in Sec. 2.3 are deterministic and so are the requirements we express in Sec. 2.4. SI designs generally do not satisfy the RIP and allow us to very finely quantize the parameter space, with spacing between delays smaller than the reciprocal of the bandwidth, spacing between Dopplers

smaller than the inverse of the signal duration, etc. This oversampling allows us to reduce the modeling errors in the linear expansion for the general class of multi-path models considered, without any increase in signal bandwidth, duration, or array elements in the system. As such, the SI requirements are rather loose when compared to those resulting from the RIP.

2.2 Sparse Mapping of MIMO Channels

Consider a MIMO system for which the narrowband approximation holds, with uniform linear arrays consisting of N_{tx} transmit elements and N_{rx} receive elements each separated by a distance d. At most Q dominant propagation paths are assumed to be present, from scatterers in the far field such that the signal from each scatterer approximately arrives as a planar wave at the receive elements. Thanks to the narrowband approximation [51], the signal arrivals at different array elements differ only by a phase rotation:

$$y_u(kT_s) = \sum_{v=0}^{N_{tx}-1} \sum_{q=1}^{Q} a_q s_v(kT_s - \tau_q) e^{j2\pi(\mu_q v + \nu_q u + k f_q)} \tag{2.1}$$

where $s_v(kT_s)$ are samples of the signal transmitted from the v-th transmitter. The parameters $\nu_q = \frac{d}{\lambda}\sin\theta_q$ and $\mu_q = -\frac{d}{\lambda}\cos\phi_q$ are the angles of arrival and departure respectively, while $f_q = \xi_q T_s$ is the normalized Doppler shift with $\xi_q \in [-\xi_{max}/2, \xi_{max}/2]$. The parameter τ_q is the relative delay of the q-th propagation path plus delay from possible synchronization errors at the receiver, i.e. $\tau_q \in [-d_\tau, \tau_{max} - d_\tau]$ where d_τ is the synchronization error; phase shifts are all included in the constant path attenuation $a_q \in \mathbb{C}$. At the u-th receiver, K samples form the vector:

$$\mathbf{y}_u = \sum_{q=1}^{Q} \mathbf{\Omega}(f_q)\mathbf{S}(\tau_q)\mathbf{e}(\mu_q)e^{j2\pi\nu_q u} a_q \tag{2.2}$$

where the $K \times K$ diagonal matrix $\{\mathbf{\Omega}(f_q)\}_{k,k} = e^{j2\pi f_q k}$. The v-th column of the $K \times N_{tx}$ matrix $\{\mathbf{S}(\tau_q)\}_{k,v} = s_v(kT_s - \tau_q)$ consists of the signal transmitted from the v-th transmitter delayed by τ_q while the vector $\mathbf{e}(\mu_q) = (1, e^{j2\pi\mu_q}, ..., e^{j2\pi\mu_q(N_{tx}-1)})^T$ is the steering vector associated with its angle of departure. The $K \times N_{rx}$ array of received vectors $\mathbf{y}_u$ is:

$$\begin{aligned} \mathbf{Y} = (\mathbf{y}_1, ..., \mathbf{y}_{N_{rx}}) &= \sum_{q=1}^{Q} \mathbf{\Omega}(f_q)\mathbf{S}(\tau_q)\mathbf{e}(\mu_q)\mathbf{e}^T(\nu_q)a_q \\ &= \mathbf{M}(\mu, f, \tau)\mathbf{A}\mathbf{M}^T(\nu) \end{aligned} \tag{2.3}$$

where $\mathbf{e}(\nu_q) = (1, e^{j2\pi\nu_q}, ..., e^{j2\pi\nu_q(N_{rx}-1)})^T$ is the steering vector associated with the angle of arrival of the q-th propagation path and the elements of the $K \times Q$ matrix $\{\mathbf{M}(\mu, f, \tau)\}_{k,q} = \{\mathbf{\Omega}(f_q)\mathbf{S}(\tau_q)\mathbf{e}(\mu_q)\}_k$. The diagonal matrix $\{\mathbf{A}\}_{q,q} = a_q$ weights each of these signals with the appropriate attenuation for this propagation path, while the $Q \times N_{tx}$ matrix $\{\mathbf{M}^T(\nu)\}_{q,u} = e^{j2\pi\nu_q u}$ is the classic Uniform Linear Array (ULA) manifold. In the model above, the input signal and the system response are blended in the term $\mathbf{M}(\mu, f, \tau)\mathbf{A}\mathbf{M}^T(\nu)$.

For the class of inputs such that $s_v(t) = \sum_{n=0}^{N-1} x_v[n]p(t - nT_s)$ where $p(t)$ is the pulse shaping filter, it is of interest to define a discrete time system response. For such inputs, $\mathbf{S}(\tau_q) = \mathbf{\Phi}(\tau_q)\mathbf{X}$ where the $K \times N$ matrix $\{\mathbf{\Phi}(\tau_q)\}_{k,n} = p((k-n)T_s - \tau_q)$ contains delayed versions of the sampled pulse and the $N \times N_{tx}$ matrix $\{\mathbf{X}\}_{n,v} = x_v[n]$. Assuming as usual that the channel is well approximated by a response with memory $L - 1$:

$$\mathbf{y} = \text{vec}(\mathbf{Y}) = \boldsymbol{\mathcal{X}}\mathbf{h} \tag{2.4}$$

where the $KN_{rx} \times LKN_{tx}N_{rx}$ matrix $\boldsymbol{\mathcal{X}}$ contains the data symbols and $\mathbf{h}$ is the $LKN_{tx}N_{rx}$ length vector of channel coefficients:

$$\begin{aligned} &\boldsymbol{\mathcal{X}}_{k+uK, l+cL+vLK+dLKN_{tx}} = x_v[k-l]\delta[c-k]\delta[d-u] \\ &\mathbf{h} = \textstyle\sum_{q=1}^{Q} \left(\mathbf{e}(\nu_q) \otimes \mathbf{e}(\mu_q) \otimes \mathbf{e}(f_q) \otimes \mathbf{p}(\tau_q)\right) a_q \end{aligned} \tag{2.5}$$

with $\{\mathbf{p}(\tau_q)\}_l = p(lT_s - \tau_q)$.

We define the vector $\mathbf{a} = (a_1, ..., a_Q)^T$ and the q-th path vector $\boldsymbol{\kappa}_q = (\nu_q, \mu_q, f_q, \tau_q)^T$. In the following proposition we develop a basis expansion for the observation $\mathbf{Y}$ and for the channel impulse response $\mathbf{h}$ by quantizing $\boldsymbol{\kappa}_q$. The basis elements are the matrices $\mathbf{M}(\mu, f, \tau)$, $\mathbf{M}^T(\nu)$, and the vectors $(\mathbf{e}(\nu) \otimes \mathbf{e}(\mu) \otimes \mathbf{e}(f) \otimes \mathbf{p}(\tau))$, evaluated at the specific quantization points.

Proposition 2.2.1. *Assume that each path parameter vector $\boldsymbol{\kappa}_q$ lies in a region of $\mathbb{R}^4$ where we place a finite grid of points $(\nu_{r_\nu}, \mu_{r_\mu}, f_{r_f}, \tau_{r_\tau})$. Each of these points is identified by an index vector $(r_\nu, r_\mu, r_f, r_\tau)$ and for each grid point we define $\boldsymbol{\Omega}_{r_f} = \boldsymbol{\Omega}(f_{r_f})$, $\mathbf{S}_{r_\tau} = \mathbf{S}(\tau_{r_\tau}), \mathbf{e}_{r_\mu} = \mathbf{e}(\mu_{r_\mu}), \mathbf{e}_{r_\nu} = \mathbf{e}(\nu_{r_\nu})$. Suppose that for every path vector $\boldsymbol{\kappa}_q$ in the actual model there exists a nearby point in the grid indexed by $(r_\nu^{(q)}, r_\mu^{(q)}, r_f^{(q)}, r_\tau^{(q)})$. The system output in* (2.3) *and the channel impulse response can be expressed respectively as:*

$$\mathbf{Y} = \mathbf{M}_{in} \boldsymbol{\mathcal{A}}^\star \mathbf{M}_{out}^T + \boldsymbol{\varepsilon}_{\mathbf{Y}}; \quad \mathbf{h} = \boldsymbol{G}\boldsymbol{\alpha}^\star + \boldsymbol{\varepsilon}_{\mathbf{h}} \tag{2.6}$$

where $\boldsymbol{\varepsilon}_{\mathbf{y}}$ and $\boldsymbol{\varepsilon}_{\mathbf{h}}$ represent the modeling error associated with approximating the true path vector on the grid. The vector $\boldsymbol{\alpha}^\star$ has elements

$$\alpha^\star_{i(\mathbf{r})+R_f R_\mu R_\tau r_\nu} = \sum_{q=1}^{Q} \delta[r_\tau - r_\tau^{(q)}]\delta[r_f - r_f^{(q)}]\delta[r_\nu - r_\nu^{(q)}]\delta[r_\mu - r_\mu^{(q)}]a_q$$

and the $K \times R_f R_\tau R_\mu$ matrix $\mathbf{M}_{in}$, the $R_f R_\tau R_\mu \times R_\nu$ matrix $\boldsymbol{\mathcal{A}}^\star$, the $R_\nu \times N_{rx}$ matrix $\mathbf{M}_{out}^T$ and the $LKN_{tx}N_{rx} \times R_f R_\mu R_\tau R_\nu$ matrix $\boldsymbol{G}$ are:

$$\{\mathbf{M}_{in}\}_{k,i(\mathbf{r})} = \{\boldsymbol{\Omega}_{r_f} \mathbf{S}_{r_\tau} \mathbf{e}_{r_\mu}\}_k, \tag{2.7}$$

$$\{\mathbf{M}_{out}^T\}_{r_\nu,u} = \{\mathbf{e}^T(r_\nu)\}_u, \tag{2.8}$$

$$\{\boldsymbol{\mathcal{A}}^\star\}_{i(\mathbf{r}),r_\nu} = \alpha^\star_{r_\mu,r_f,r_\tau,r_\nu} \tag{2.9}$$

$$\{\boldsymbol{G}\}_{i(\mathbf{r})+R_f R_\mu R_\tau r_\nu} = \mathbf{e}_{r_\nu} \otimes \mathbf{e}_{r_\mu} \otimes \mathbf{e}_{r_f} \otimes \mathbf{p}_{r_\tau} \tag{2.10}$$

with $\mathbf{r} = (r_\mu, r_f, r_\tau)$ and the index $i(\mathbf{r}) = r_\tau + r_f R_\tau + r_\mu R_\tau R_f$.

Proof. The proof is obtained observing that the sum over Q terms in (2.3) and (2.5), for $\boldsymbol{\kappa}_q, q = 1, \ldots, Q$ on the grid point indexed by $(r_\nu^{(q)}, r_\mu^{(q)}, r_f^{(q)}, r_\tau^{(q)})$, can be replaced with a sum of indicator functions $\sum_{q=1}^{Q} \delta[r_\tau - r_\tau^{(q)}]\delta[r_f - r_f^{(q)}]\delta[r_\nu - r_\nu^{(q)}]\delta[r_\mu - r_\mu^{(q)}]a_q$ that single out the existing elements in the summation, while setting all other coefficients to zero. □

Under the assumptions in Prop. 2.2.1, we have an equivalent vector model:

$$\mathbf{y} = \mathrm{vec}(\mathbf{Y}) = \underbrace{(\mathbf{M}_{out} \otimes \mathbf{M}_{in})}_{\mathbf{M}} \boldsymbol{\alpha}^\star + \boldsymbol{\varepsilon}_\mathbf{y} \tag{2.11}$$

Definition 2.2.2. *The system is said to be sparse if the number of non-zero coefficients in $\boldsymbol{\alpha}^\star \in \mathbb{C}^R$ (or cardinality of the support of $\boldsymbol{\alpha}^\star$), with $R = R_f R_\mu R_\tau R_\nu$, is less than half its length. We denote by S the cardinality of the support of $\boldsymbol{\alpha}^\star$, i.e. $S = \|\boldsymbol{\alpha}^\star\|_0$.*

The expansion in Prop. 2.2.1 is exact if the $\boldsymbol{\kappa}_q$ fall on the chosen grid in $\mathbb{R}^4$ and thus $\boldsymbol{\varepsilon}_\mathbf{y} = \boldsymbol{\varepsilon}_\mathbf{h} = \mathbf{0}$. As the level of acceptable modeling error in (2.6) increases, the class of approximations widens, including sparse vectors whose support S can come close to Q, possibly even smaller. Since a possible approximation is given by the Taylor expansion of the path vectors $\boldsymbol{\kappa}_q$ around closest grid point, we know there exist approximating vectors with sparsity $S \approx Q$ that result in vanishing modeling error as the quantization grid is made finer. One should notice that in the absence of noise any model that gives zero modeling error on the impulse response, if it can be retrieved from the observations, is best, no matter how sparse it is. However, using a parsimonious model is effective in reducing the error in the presence of noise. Reducing the ambiguity intrinsic in the problem is a necessary first step to find designs that remove undesired spurious solutions fitting the observations while not fitting the channel. This motivates us to look into the subclass of problems that

are sparse with zero modeling error and examine the effects of modeling errors on the continuous parameters numerically. As discussed in the next section, SI inputs are those inputs $s_v(t)$ that allow us to identify $\boldsymbol{\alpha}^\star$ and are sufficient to construct $\mathbf{h}$.

2.3 Sufficiently Informative Inputs and Identification of Sparse Channels

At the receiver, the observations $\mathbf{y}$ are projected onto the sub-space corresponding to the known pilot signal:

$$\tilde{\mathbf{y}} = \underbrace{(\mathbf{I}_{N_{rx}} \otimes \tilde{\mathbf{P}})}_{\mathbf{P}} \mathbf{y} = \underbrace{\mathbf{P}\mathbf{M}}_{\boldsymbol{\Upsilon}} \boldsymbol{\alpha}^\star + \boldsymbol{\varepsilon}_{\tilde{\mathbf{y}}} \tag{2.12}$$

where $\mathbf{P}$ is the projection matrix. For example, in a SISO OFDM system, the matrix $\mathbf{P}$ is an $M \times K$ DFT matrix where K is the number of subcarriers and M the number of pilots. Our aim is to recover the channel impulse response via the model (2.12).

In this section we assume that for a specified grid there exists a sparse vector $\boldsymbol{\alpha}^\star$ such that the resulting $\boldsymbol{\varepsilon}_{\tilde{\mathbf{y}}} = \boldsymbol{\varepsilon}_{\mathbf{h}} = \mathbf{0}$. In general, assuming $\boldsymbol{G}$ is full row rank and fat and thus has as a non-trivial null-space, there exist infinite solutions to $\boldsymbol{G}\boldsymbol{\alpha} = \mathbf{h}$, irrespective of the grid choice and true parameters. There are two types of errors that may result from our model. The first is the observation reconstruction error $\boldsymbol{\varepsilon}_{\tilde{\mathbf{y}}} \triangleq \tilde{\mathbf{y}} - \boldsymbol{\Upsilon}\boldsymbol{\alpha}$, while the second is the channel reconstruction error $\boldsymbol{\varepsilon}_{\mathbf{h}} \triangleq \mathbf{h} - \boldsymbol{G}\boldsymbol{\alpha}$. If our goal is to estimate $\mathbf{h}$, what matters is the residual $\boldsymbol{\varepsilon}_{\mathbf{h}}$ associated with the solution. We note that a solution $\boldsymbol{\alpha}$ such that $\boldsymbol{\varepsilon}_{\tilde{\mathbf{y}}} = \mathbf{0}$ does not imply $\boldsymbol{\varepsilon}_{\mathbf{h}} = \mathbf{0}$, since the matrix $\mathbf{P}\boldsymbol{\mathcal{X}}$ may have a non-trivial null-space. Our goal is to find $\boldsymbol{\alpha}^\star$ from

the samples $\tilde{\mathbf{y}}$ by exploiting the assumed sparsity. One may think that this effort is futile since in the absence of noise all we desire is the smallest $\boldsymbol{\varepsilon}_{\mathbf{h}}$ possible, i.e. $\boldsymbol{\varepsilon}_{\mathbf{h}} = \mathbf{0}$. However, methods to retrieve sparse solutions are desirable because they reduce the model order, leading to a better tradeoff between modeling error and noise error. A first step in obtaining a working method is to ensure that if a sparse model exists, it is observable through the vector $\tilde{\mathbf{y}}$ associated with the training signal. In the absence of noise, the criterion we use to find the sparse solution $\boldsymbol{\alpha}^\star$ is the following optimization:

$$\hat{\boldsymbol{\alpha}} = \arg\min \|\boldsymbol{\alpha}\|_0 \;\text{ s.t. }\; \tilde{\mathbf{y}} = \boldsymbol{\Upsilon}\boldsymbol{\alpha}. \tag{2.13}$$

The following lemma provides conditions such that the solution $\hat{\boldsymbol{\alpha}}$ to (2.13) is the vector $\boldsymbol{\alpha}^\star$ defined in Proposition 2.2.1.

Lemma 2.3.1. *Assuming Prop. 2.2.1 holds with $\boldsymbol{\varepsilon}_{\mathbf{h}} = \mathbf{0}0$ and $\boldsymbol{\alpha}^\star$ is sparse, i.e. $\|\boldsymbol{\alpha}^\star\|_0 \leq S \leq R/2$ and R is the number of columns in $\boldsymbol{\Upsilon}$, in order for the solution to* (2.13) *be unique and equal to $\boldsymbol{\alpha}^\star$, every subset of $2S$ columns of $\boldsymbol{\Upsilon}$ must be linearly independent.*

Proof. See Appendix A.1. □

Note that the Kruskal rank [52] of a matrix is the largest value m such that every subset of m columns of the matrix is linearly independent. Thus the condition in Lemma 2.3.1 is equivalent to saying that the Kruskal rank of the matrix $\boldsymbol{\Upsilon}$ must be greater than $2S$.

The assumption that the parameters $\boldsymbol{\kappa}_q$ always fall on the constructed grid is impossible to maintain in practice, however we have assumed there exists a sparse solution $\boldsymbol{\alpha}^\star$ with $\|\boldsymbol{\alpha}^\star\|_0 \leq S$ and $\boldsymbol{\varepsilon}_{\mathbf{h}} = \mathbf{0}$. This restricted set we analyze is a subset

of the original set with continuous parameters, but it is not easy to conclude that this set contains only channels whose path parameters fall on the grid as we cannot exclude cases where $\|\boldsymbol{\alpha}\|_0 \leq S$ exist, with $S \leq Q$ and $S \leq R/2$ and having zero approximation error on the grid. Under the assumption a sparse solution exists, we define a sufficiently informative input as follows:

Definition 2.3.2. *Assuming for a predefined basis* $\boldsymbol{G}$*, the parameters* $\boldsymbol{\kappa}_q$*,* $q = 1, \ldots, Q$ *are such that there is a unique sparsest solution to* $\mathbf{h} = \boldsymbol{G}\boldsymbol{\alpha}^\star$ *with* $\|\boldsymbol{\alpha}^\star\|_0 \leq S$*,* ***Sufficiently informative inputs*** *are those input signals* $s_v(t)$ *such that* $\boldsymbol{\alpha}^\star$ *is also the unique solution to the optimization* (2.13)*.*

In the following we consider the design of SI inputs $s_v(t)$, in some cases also placing constraints on the grid for the parameters $\boldsymbol{\kappa}_q$, such that we can identify $\boldsymbol{\alpha}^\star$ provided we have an upper bound on number of nonzero elements S. The intuition behind our designs is that if different $\boldsymbol{\kappa}_q$ on the grid correspond to linearly independent vectors having a Vandermonde structure, and the number of observation is large enough, the input should be SI as will be made clear in Sec. 2.4. We reiterate that this does not establish identifiability of all $\mathbf{h}$ as it is impossible to avoid channel modeling error when choosing the discrete grid for the parameter space. It does, however, eliminate structural ambiguities in the parameter search that is implicitly performed when searching the support of $\boldsymbol{\alpha}$.

2.3.1 Restricted Isometry Property and Mutual Coherence

The optimization (2.13) is in general a combinatorial problem. There have been a number of methods proposed to approximate this problem, including greedy algorithms such as matching pursuit [64], iteratively re-weighted least squares methods

such as FOCUSS [41], penalized least squares problems such as LASSO and Basis Pursuit [89], [24], and Bayesian methods [90], [75]. In [24], [21] a tractable relaxation of (2.13) was examined, called *Basis Pursuit*:

$$\hat{\boldsymbol{\alpha}} = \arg\min \|\boldsymbol{\alpha}\|_1 \;\; \text{s.t.} \;\; \tilde{\mathbf{y}} = \boldsymbol{\Upsilon}\boldsymbol{\alpha} \tag{2.14}$$

This is a convex optimization problem and can be solved using standard convex optimization algorithms. The uniqueness of the solution of (2.14) is tied to the following: [20], [28]:

Restricted Isometry Property (RIP) ([20]): *Assume the columns of* $\boldsymbol{\Upsilon}$ *have unit Euclidean norm. If for every subset of* $T \leq S$ *columns of* $\boldsymbol{\Upsilon}$*, the matrix* $\boldsymbol{\Upsilon}_{\mathcal{T}}$ *containing them is such that*

$$(1-\delta_S)\|\mathbf{v}\|_2^2 \leq \|\boldsymbol{\Upsilon}_{\mathcal{T}}\mathbf{v}\|_2^2 \leq (1+\delta_S)\|\mathbf{v}\|_2^2$$

with $0 < \delta_S < 1$*, the matrix is said to satisfy the* S*-restricted isometry property with restricted isometry constant* δ_S*.*

Mutual Coherence ([28]): *Assume the columns of* $\boldsymbol{\Upsilon}$ *have unit Euclidean norm. The mutual coherence* $\mu_c(\boldsymbol{\Upsilon}) \triangleq \max_{i \neq j} |\{\boldsymbol{\Upsilon}^H\boldsymbol{\Upsilon}\}_{i,j}|$.

Two different sufficient conditions for equivalence of the solutions to (2.13) and (2.14) based on these metrics are given in the following lemma:

Lemma 2.3.3 ([22], [28])**.** *The solution to the problem* (2.14) *is unique and equivalent to the solution of* (2.13) *for any* $\boldsymbol{\alpha}^\star$ *such that* $\|\boldsymbol{\alpha}^\star\|_0 \leq S$ *if either of the following two sufficient conditions are satisfied:*

$$\delta_S + \delta_{2S} + \delta_{3S} < 1 \;\; \textit{or} \;\; \mu_c(\boldsymbol{\Upsilon}) < \frac{1}{2S-1}$$

In general, the sufficient conditions in Lemma 2.3.3 (and other similar conditions) are pessimistic. It well known that cases exist where neither of the sufficient

conditions in Lemma 2.3.3 are satisfied but the matrix $\boldsymbol{\Upsilon}$ It is for this reason, when considering pilot signal designs in Sec. 2.4, that we focus on designs guaranteeing identifiability using (2.13), namely every subset of $2S$ columns of $\boldsymbol{\Upsilon}$ is linearly independent, implying a unique sparsest solution and preventing spurious solutions of the problem (2.13) from existing. This is precisely the meaning of a SI input in [61]. The results presented here and in Sec. 2.4 apply only in the absence of noise and assume that a sparse solution $\boldsymbol{\alpha}^\star$ exists such that $\boldsymbol{\varepsilon}_\mathbf{h} = \mathbf{0}$. In chapter 3 we discuss how the basic guidelines provided by Def. 2.3.2 can be used, with additional constraints, to provide robust designs when noise and modeling error are present in the system.

Note the condition on the linear independence of subsets of columns of $\boldsymbol{\Upsilon}$ is equivalent to guaranteeing the minimum eigenvalue of $\boldsymbol{\Upsilon}_\mathcal{T}^H \boldsymbol{\Upsilon}_\mathcal{T}$ is greater than zero over all subsets $\mathcal{T}$ of size $|\mathcal{T}| \leq 2S$. This is equivalent to the *Unique Reconstruction Property* in [41].

2.4 Sufficiently Informative Training Designs

The Vandemonde matrix is frequently encountered in both engineering and mathematics and has many useful properties [46,73]. In the following, assuming a sparse solution $\|\boldsymbol{\alpha}^\star\|_0 \leq S$ exists with $\boldsymbol{\varepsilon}_\mathbf{h} = 0$, our strategy will be to construct inputs such that the noise-free observations may be mapped via projections[2] (2.12) to sparse linear combinations of S columns from a Vandermonde matrix with the following form:

$$\{\boldsymbol{\Upsilon}\}_{m,r} = \upsilon_r^m, \quad m = 0, ..M-1 \ , r = 0, .., R-1 \tag{2.15}$$

[2]The projection matrices $\mathbf{P}$ corresponding to each mapping may be found in the proofs.

We will use the following lemma, a well known result [46], though we state it for clarity:

Lemma 2.4.1. *If the parameters* $(v_0^1, ...v_{R-1}^1)$ *of the* $M \times R$ *matrix* $\boldsymbol{\Upsilon}$ *are distinct, then any* $M \leq R$ *columns of* $\boldsymbol{\Upsilon}$ *are linearly independent.*

Thus if our observations are mapped to a sparse linear combination of $S \leq R/2$ columns (see Def. 2.2.2) from a matrix $\boldsymbol{\Upsilon}$ satisfying the conditions of this lemma with $M \geq 2S$, any $2S$ columns are linearly independent and as a result the conditions in Lemma 2.3.1 are satisfied. In some cases our observations may be mapped into a subset of rows and columns from a Vandermonde matrix and we will make use of Chebotarev's Theorem [83, 84]:

Lemma 2.4.2. *For an* $M \times R$ *matrix* $\boldsymbol{\Upsilon}$ *with entries:*

$$\{\boldsymbol{\Upsilon}\}_{m,r} = e^{j2\pi c_m r/P}, \quad c_m \in \{0, ..., K-1\}, \ r = 0, ..., R-1,$$

if P *is prime,* $P \geq R \geq K$*, and the* $M \leq K$ *elements* c_m *are distinct, any* M *columns of* $\boldsymbol{\Upsilon}$ *are linearly independent.*

This lemma was used to prove Lemma 2.1.2 in the introduction. It implies that any square submatrix of the $P \times P$ DFT matrix, with prime P, is full rank, and any tall submatrix has full column rank. Thus, for prime $P \geq R \geq M \geq 2S$, when our observations are mapped to a sparse linear combination of $S \leq R/2$ columns of $\boldsymbol{\Upsilon}$ having the form in this lemma, any $2S$ columns of this matrix are linearly independent, i.e. no null vector can have support smaller or equal to $2S$. Next, we provide SI input designs for sparse models by mapping our observations to sparse linear combinations of columns from $\boldsymbol{\Upsilon}$.

2.4.1 SISO Channels

In all the SISO channels $N_{tx} = N_{rx} = 1$, so that $\mathbf{e}_{r_\nu} = \mathbf{e}_{r_\mu}^T = 1$ and $\mathbf{S}_{r_\tau} = \mathbf{s}_{r_\tau}$ in (2.11):

$$\mathbf{y} = \sum_{\mathbf{r}=(r_\tau, r_f)} \mathbf{\Omega}_{r_f} \mathbf{s}_{r_\tau} \{\boldsymbol{\alpha}\}_{i(\mathbf{r})}, \quad \mathbf{y} \in \mathbb{C}^K \tag{2.16}$$

SISO Time-Selective Channel

We begin with a simple and intuitive example, the time-varying flat-fading channel. As there are no delays, the vector $\mathbf{s}_{r_\tau} = \mathbf{s}$ in (2.16) where $\{\mathbf{s}\}_k = s(kT_s)$. Over a window of K observations, the vector $\mathbf{s}$ contains M pilot symbols at locations $c_m \in \{0, .., K-1\}$. Collecting K observations:

$$\mathbf{y} = \sum_{r_f=0}^{R_f-1} \mathbf{\Omega}_{r_f} \mathbf{s} \alpha_{r_f} \tag{2.17}$$

The following theorem provides a SI input design for this system.

Theorem 2.4.3. *In time-selective SISO systems that are sparse over a grid $f_{r_f} = r_f \epsilon_f - d_f$ with $\epsilon_f R_f < 1$ and have a unique sparsest representation $\mathbf{y} = \mathbf{M}\boldsymbol{\alpha}^\star$ with $\|\boldsymbol{\alpha}^\star\|_0 \leq S$, identifiability of the channel is guaranteed when the pilot signal $s(t)$ consists of $M \geq 2S$ non-zero samples and $1/\epsilon_f$ prime.*

Proof. See Appendix A.2 □

If the received signal consists of S Doppler shifted versions of the transmitted signal, an input with at least $2S$ nonzero pilot symbols $s(c_m T_s)$ is sufficiently informative. In the proof of Theorem 2.4.3, the observations are mapped to linear combinations of columns from the $M \times R_f$ matrix $\{\mathbf{\Upsilon}\}_{m,r_f} = e^{j2\pi c_m f_{r_f}}$. When

the pilot symbols are uniformly spaced ($c_m = mN_s$), $\boldsymbol{\Upsilon}$ has the form in Lemma 2.4.1, and the assumption of a uniform frequency grid (i.e $f_{r_f} = r_f\epsilon_f - d_f$ with spacing ε_f and shift d_f) and prime $1/\epsilon_f$ is not necessary, only that the elements $e^{j2\pi N_s f_{r_f}}$ are unique and $M \geq 2S$. We remark that due to the sampled nature of the received signal, in order to satisfy the conditions of Lemma 2.3.1, it is necessary for a sufficiently informative input to contain at least $2S$ non-zero samples.

SISO Frequency-Selective Channel

The time-invariant frequency-selective channel is the dual of the time-selective case. Assuming no Doppler shifts, the matrix $\boldsymbol{\Omega}_{r_f} = \mathbf{I}$ in (2.16). Consider the transmission of a pilot signal $s(t) = \frac{1}{\sqrt{K}} \sum_{m=0}^{M-1} \tilde{s}_m e^{-j2\pi \frac{c_m t}{KT_s}}$ with $c_m \in \{0, ..., K-1\}$ and $M \leq K$. We assume $s(t)$ contains a cyclic prefix of $N_{cp}T_s \geq \tau_{\max}$. Collecting K samples after removing the cyclic prefix:

$$\mathbf{y} = \sum_{r_\tau=0}^{R_\tau-1} \mathbf{s}_{r_\tau} \alpha_{r_\tau} \tag{2.18}$$

The following theorem provides a SI input design for this system.

Theorem 2.4.4. *In frequency selective SISO systems that are sparse over a grid $\tau_{r_\tau} = r_\tau\epsilon_\tau - d_\tau$ with $\epsilon_\tau R_\tau \leq KT_s$, and have a unique sparsest representation $\mathbf{y} = \mathbf{M}\boldsymbol{\alpha}^\star$ with $\|\boldsymbol{\alpha}^\star\|_0 \leq S$, identifiability of the channel is guaranteed when the pilot signal $s(t)$ consists of a weighted sum of $M \geq 2S$ pilot tones $s(t) = \sum_{m=0}^{M-1} \tilde{s}_m e^{-j2\pi \frac{c_m}{KT_s} t}$ and KT_s/ϵ_τ prime.*

Proof. See Appendix A.3 □

In the proof, the observations are mapped to linear combinations of columns from the $M \times R_\tau$ matrix $\{\boldsymbol{\Upsilon}\}_{m,r_\tau} = e^{j2\pi c_m \frac{\tau_{r_\tau}}{KT_s}}$. As in the time-selective case,

the assumption of a uniform delay grid $\tau_{r_\tau} = r_\tau \epsilon_\tau - d_\tau$ is not needed when the pilot tones are uniformly spaced, i.e. $c_m = mN_s$, only that elements $e^{j2\pi N_s \frac{\tau_{r_\tau}}{KT_s}}$ are distinct and $M \geq 2S$.

Cyclic-prefixed pilot signals can always be expressed via a Fourier series. As in the time-selective case, we remark that in order for the class of cyclic-prefixed pilot signals to be sufficiently informative, it is necessary that at least $M \geq 2S$ spectral components be non-zero in order to satisfy the conditions in Lemma 2.3.1.

SISO Doubly-Selective Channel

We observed for the time-selective channels, an input with a certain amount of temporal support was SI and likewise one with a certain amount of spectral support was SI for frequency-selective channels. For the doubly-selective channel, which contains both Doppler shifts as well as delays, we propose the use of a chirped (or linear frequency modulated) pilot signal for identification i.e. $s(t) = e^{-j\pi \frac{\beta}{KT_s^2} t^2}$ for $-\tau_{max} \leq t < KT_s$. Taking K samples after removing a portion τ_{max} corresponding to the prefix:

$$\mathbf{y} = \sum_{\mathbf{r}=(r_f, r_\tau)} \mathbf{\Omega}_{r_f} \mathbf{s}_{r_\tau} \{\boldsymbol{\alpha}\}_{i(\mathbf{r})} \tag{2.19}$$

The following theorem provides a SI input for this system.

Theorem 2.4.5. *In doubly-selective SISO systems that have a unique sparsest representation* $\mathbf{y} = \mathbf{M}\boldsymbol{\alpha}^\star$ *with* $\|\boldsymbol{\alpha}^\star\|_0 \leq S$ *and pilot signal* $s(t) = e^{-j\pi \frac{\beta}{KT_s^2} t^2}$*, identifiability of the channel is guaranteed if the elements* $e^{j2\pi(f_{r_f} + \frac{\beta}{KT_s}\tau_{r_\tau})}$ *are unique and* $K \geq 2S$.

Proof. See Appendix A.4 □

In the proof, the observations $\mathbf{y}$ are mapped into linear combinations of columns from a $K \times R_f R_\tau$ matrix $\{\boldsymbol{\Upsilon}\}_{k,r_\tau+R_\tau r_f} = e^{j2\pi k(f_{r_f}+\frac{\beta}{KT_s}\tau_{r_\tau})}$ having the form in Lemma 2.4.1 (with $M = K$) and thus the condition that each of the elements $e^{j2\pi(f_{r_f}+\frac{\beta}{KT_s}\tau_{r_\tau})}$ is distinct. Note that the chirped pilot signal is SI for the time-varying SISO channel (or frequency-selective SISO channel) where the elements $e^{j2\pi f_{r_f}}$ (or $e^{j2\pi\frac{\beta}{KT_s}\tau_{r_\tau}}$ for the frequency-selective SISO channel) must be unique.

2.4.2 MISO Channels

In MISO systems $N_{rx} = 1$, implying the vector $\mathbf{e}_{r_\nu} = 1$ in (2.11). In this case:

$$\mathbf{y} = \sum_{\mathbf{r}=(r_\mu,r_f,r_\tau)} \boldsymbol{\Omega}_{r_f}\mathbf{S}_{r_\tau}\mathbf{e}_{r_\mu}\{\boldsymbol{\alpha}\}_{i(\mathbf{r})}, \quad \mathbf{y} \in \mathbb{C}^K \tag{2.20}$$

Next, we examine the case of the MISO time-selective and frequency-selective systems.

MISO Time-Selective Channel

As there are no delays, the matrix $\mathbf{S}_{r_\tau} = \mathbf{S}$ in (2.20). We observed in the SISO time-selective system that an input with a certain amount of temporal support was SI. Applying the same reasoning, we transmit a signal $s_v(t) = \tilde{s}_v w_v(t)$ from each transmitter, where $w_v(t)$ is a rectangular windowing function that is non-zero only for $vT_s \leq t < (v+1)T_s$, implying that $\mathbf{S}$ is a diagonal matrix $\{\mathbf{S}\}_{v,v} = \tilde{s}_v$. Collecting K observations:

$$\mathbf{y} = \sum_{\mathbf{r}=(r_\mu,r_f)} \boldsymbol{\Omega}_{r_f}\mathbf{S}\mathbf{e}_{r_\mu}\{\boldsymbol{\alpha}\}_{i(\mathbf{r})} \tag{2.21}$$

The following theorem provides a SI input design for this system.

Theorem 2.4.6. *In time-varying MISO systems that have a unique sparsest representation* $\mathbf{y} = \mathbf{M}\boldsymbol{\alpha}^\star$ *with* $\|\boldsymbol{\alpha}^\star\|_0 \leq S$*, when the pilot signal from the* v*-th transmit antenna* $s_v(t) = s(t)w_v(t)$*, identifiability of the channel is guaranteed if the elements* $e^{j2\pi(f_{r_f}+\mu_{r_\mu})}$ *are unique and the number of transmit antennas* $N_{tx} \geq 2S$*.*

Proof. See Appendix A.5. □

The observations are mapped in the proof to linear combinations of columns from the $K \times R_f R_\mu$ matrix $\{\boldsymbol{\Upsilon}\}_{k,r_f+r_\mu R_f} = e^{j2\pi k(f_{r_f}+\mu_{r_\mu})}$ having the form in Lemma 2.4.1 , with $M = K$. The condition for identifiability results by guaranteeing the $R_f R_\mu$ elements $e^{j2\pi(f_{r_f}+\mu_{r_\mu})}$ are unique. For this SI input design, each sample of $\mathbf{y}$ corresponds to a different transmit antenna since we send one pilot signal from each antenna sequentially in time (i.e. $M = N_{tx}$). In Theorem 2.4.3 we observed that an input with at least $2S$ pilot symbols was SI for time-varying channels and thus the condition $N_{tx} \geq 2S$. If no Doppler shifts were present, we could identify the angles of arrival in the same manner with $N_{tx} \geq 2S$. The additional assumption of sparsity allows us to identify the time-varying MISO channel with only $N_{tx} \geq 2S$.

MISO Frequency-Selective Channel

In the frequency-selective system we assume no Doppler shifts, implying $\boldsymbol{\Omega}_{r_f} = \mathbf{I}$ in (2.20). For the SISO frequency-selective system an input with a certain amount amount of spectral support was SI, and thus we will transmit a single tone from each transmit antenna. Let the pilot signal on the v-th antenna $s_v(t) = e^{-j2\pi vt/(KT_s)}$ for $-\tau_{max} \leq t \leq KT_s$ such that these frequencies are orthogonal over K observa-

tions. Collecting K observations after removing the cyclic prefix:

$$\mathbf{y} = \sum_{\mathbf{r}=(r_\mu, r_\tau)} \mathbf{S}_{r_\tau} \mathbf{e}_{r_\mu} \{\boldsymbol{\alpha}\}_{i(\mathbf{r})} \tag{2.22}$$

The following theorem provides a SI input design for this system.

Theorem 2.4.7. *In frequency-selective MISO systems that have a unique sparsest representation* $\mathbf{y} = \mathbf{M}\boldsymbol{\alpha}^\star$ *with* $\|\boldsymbol{\alpha}^\star\|_0 \leq S$*, when the pilot signal from the* v*-th transmit antenna* $s_v(t) = e^{-j2\pi vt/KT_s}$*, identifiability of the channel is guaranteed if the elements* $e^{j2\pi(\mu_{r_\mu}+\tau_{r_\tau}/(KT_s))}$ *are unique and the number of transmit antennas* $N_{tx} \geq 2S$*.*

Proof. See Appendix A.6. □

In the proof, the observations are mapped to linear combinations of columns from the $N_{tx} \times R_\tau R_\mu$ matrix $\{\boldsymbol{\Upsilon}\}_{v,r_\tau+r_\mu R_\tau} = e^{j2\pi v(\mu_{r_\mu}+\tau_{r_\tau}/(KT_s))}$ having the form in Lemma 2.4.1 (with $M = N_{tx}$). As in the MISO time-selective system, the condition $N_{tx} \geq 2S$ guarantees identifiability of the angles of departure, while the number of pilot tones sufficient to identify the delays is spread over the transmit antennas.

2.4.3 Multiple Output Channels

Recall that in the multiple output systems, via Prop. 2.2.1, our noise-free observations can be expressed as $\mathbf{Y} = \mathbf{M}_{in}\boldsymbol{\mathcal{A}}\mathbf{M}_{out}^T$. In these systems, we will first identify the angle of arrival associated with each scatterer. Note that the (r_ν, k)-th element of the matrix $\tilde{\boldsymbol{\mathcal{A}}} \triangleq \boldsymbol{\mathcal{A}}^T\mathbf{M}_{in}^T$ corresponds to the k-th sample of a signal arriving at the receiver array from a propagation path with the arrival angle associated with

ν_{r_ν}. Assuming S angles, at most S rows of $\tilde{\boldsymbol{\mathcal{A}}}$ have nonzero elements. Examining the k-th snapshot at the receiver array, corresponding to the k-th column $[\mathbf{Y}^T]_k = \mathbf{M}_{out}[\tilde{\boldsymbol{\mathcal{A}}}]_k$, we arrive with the following theorem on the identifiability of the arrival angles:

Theorem 2.4.8. *For multiple output systems (SIMO and MIMO) that over a single receiver snapshot have a unique sparsest representation* $[\mathbf{Y}^T]_k = \mathbf{M}_{out}[\tilde{\boldsymbol{\mathcal{A}}}^\star]_k$ *with* $\|[\tilde{\boldsymbol{\mathcal{A}}}^\star]_k\|_0 \leq S$, $[\tilde{\boldsymbol{\mathcal{A}}}^\star]_k$ *is identifiable if* $N_{rx} \geq 2S$ *and the elements* $e^{j2\pi\nu_{r_\nu}}$ *are unique.*

Proof. See Appendix A.7. □

In Theorem 2.4.8, we are only identifying the angles of arrival and thus have no constraint on the input signal other than the signal arriving via each propagation path is present in the k-th snapshot. In other words, the theorem only places constraints on the number of receive elements sufficient to identify the angles of arrival, while the input signal need not be known to identify the angles of arrival. Further, with the addition of multiple snapshots, assuming the arrival angles do not change over these snapshots, having twice as many receive antennas as propagation paths is still a sufficient condition. This is because the multiple snapshots can be expressed as $\mathbf{Y}^T = \mathbf{M}_{out}\tilde{\boldsymbol{\mathcal{A}}}^\star$ where at most S rows in $\tilde{\boldsymbol{\mathcal{A}}}^\star$ contain non-zero elements, since the spatial sparsity pattern remains unchanged. Similar to [63] the problem (2.13), applied to the identification of the arrival angles, can be modified to accommodate multiple snapshots:

$$\hat{\boldsymbol{\mathcal{A}}} = \arg\min \|\mathrm{diag}(\tilde{\boldsymbol{\mathcal{A}}}\tilde{\boldsymbol{\mathcal{A}}}^H)\|_0 \quad \text{s.t.} \quad \mathbf{Y}^T = \mathbf{M}_{out}\tilde{\boldsymbol{\mathcal{A}}} \tag{2.23}$$

After identifying each of the S angles of arrival and applying the projection

(2.12), and given our definition of $\mathbf{M}_{in}$ and $\mathbf{Y}$, we obtain the following theorem on the identifiability of the SIMO/MIMO channels:

Theorem 2.4.9. *For multiple output systems satisfying the conditions in Theorem 2.4.8 with S denoting the number of propagation paths, the channel is uniquely identifiable if every pair of columns of $\mathbf{M}_{in}$ is linearly independent and $K \geq 2$.*

Proof. See Appendix A.8 □

Note that this theorem applies to all the cases considered in Sec.'s 2.4.1 and 2.4.2 when multiple receive antennas are present, as they all contain special cases of the matrix $\tilde{\mathbf{P}}\mathbf{M}_{in}$. In summary, when the conditions in Theorem 2.4.8 are satisfied, multiple outputs allow us to identify the arrival angle of each propagation path via (2.23). Once the angles are identified, we steer the receive array in the direction of each arrival, isolating the components of that particular propagation path, greatly reducing the requirements on the pilot design needed to identify the channel, since we are looking for only one set of parameters (μ, f, τ) associated with each arrival angle. Finding this component is as easy as determining the unit normalized column of $\tilde{\mathbf{P}}\mathbf{M}_{in}$ most correlated with the observations on this propagation path. One may wonder if a similar approach can be used for multiple input channels, however the difficulty in applying such a technique is that the angles of departure are not known apriori at the transmitters, and thus we can not beamform in the direction of each scatterer.

We remark that the angles of arrival could have alternatively been found using common methods found in the array processing literature, such as MUSIC and ESPRIT, however, as demonstrated in [12,13,63], methods using the sparsity of the system and a fine grid tend to have better performance compared to the common

subspace based methods, especially in the presence of noise.

2.5 Effects of Noise and Modeling Error

In this chapter we have examined the design of inputs that allowed, in the absence of noise, the identification of a subset of channels corresponding to sparse $\boldsymbol{\alpha}^\star$ and having channel reconstruction error $\boldsymbol{\varepsilon}_{\mathbf{h}} = \mathbf{h} - \mathbf{G}\boldsymbol{\alpha}^\star = \mathbf{0}$. As discussed in Sec. 2.3, a solution $\boldsymbol{\alpha}$ such that $\boldsymbol{\varepsilon}_{\tilde{\mathbf{y}}} = \tilde{\mathbf{y}} - \boldsymbol{\Upsilon}\boldsymbol{\alpha} = \mathbf{0}$ does not imply $\boldsymbol{\varepsilon}_{\mathbf{h}} = 0$. The ingredients of the training designs provided in Sec. 2.4 consisted of a certain quantization of the channel parameter space and an input signal together guaranteeing that $\boldsymbol{\alpha}$ solving[3] (2.13) (with $\boldsymbol{\varepsilon}_{\tilde{\mathbf{y}}} = \mathbf{0}$) implies $\boldsymbol{\varepsilon}_{\mathbf{h}} = \mathbf{0}$. We called these inputs SI, however, we remark that they are SI with respect to specific families of grid points.

With respect to the models in Sec. 2.2, in the presence of noise and modeling errors:

$$\tilde{\mathbf{y}} = \boldsymbol{\Upsilon}\boldsymbol{\alpha}^\star + \mathbf{w} + \boldsymbol{\varepsilon}_{\tilde{\mathbf{y}}}, \qquad \mathbf{h} = \boldsymbol{G}\boldsymbol{\alpha}^\star + \boldsymbol{\varepsilon}_{\mathbf{h}} \tag{2.24}$$

and $\boldsymbol{\alpha}^\star$ is the one resulting from Proposition 2.2.1, obtained by quantizing the true parameters with the grid points and retaining all (or a subset) of the columns. As discussed in Sec. 2.2, as the quantization grid is made very fine, there exist sparse $\boldsymbol{\alpha}^\star$ with vanishing modeling error $\boldsymbol{\varepsilon}_{\mathbf{h}}$. It is important to note that $\boldsymbol{\varepsilon}_{\tilde{\mathbf{y}}}$ is not a function of noise, but simply due to the modeling of the system. We will consider systems for which $\boldsymbol{\alpha}^\star$ is not only sparse, but its corresponding $\boldsymbol{\varepsilon}_{\mathbf{h}}$ is such that for any $\|\boldsymbol{\alpha}\|_0 \leq \|\boldsymbol{\alpha}^{star}\|_0$, the resulting error on the channel impulse response $\|\mathbf{h} - \boldsymbol{G}\boldsymbol{\alpha}\|^2 > \|\boldsymbol{\varepsilon}_{\mathbf{h}}\|^2$. In fact, to cope with noise, sparse recovery algorithms often

[3]We remark that due to the complex exponential structure of the matrices $\boldsymbol{\Upsilon}$ we obtained, if $\|\boldsymbol{\alpha}^\star\|_0 \leq S$ is sparse as well as *nonnegative*, it is also the unique solution to (2.14) via Corollary 1.2 in [30].

replace the equality constraint in 2.13 with an inequality $\|\tilde{\mathbf{y}} - \boldsymbol{\Upsilon}\boldsymbol{\alpha}\|_2 < \eta$, where η is a likely estimate of $\|\mathbf{w} + \boldsymbol{\varepsilon}_{\tilde{\mathbf{y}}}\|_2$. Hence, it is no longer sufficient to just control the null-space of $\boldsymbol{\Upsilon}$. Unfortunately, SI inputs only control the set of spurious solutions that arise due to the null space of $\boldsymbol{\Upsilon}$. Eventually, as η grows, it becomes impossible to prevent all spurious solutions of $\|\tilde{\mathbf{y}} - \boldsymbol{\Upsilon}\boldsymbol{\alpha}\|_2^2 < \eta$ from being as sparse (or sparser) than $\boldsymbol{\alpha}^\star$. .

When using the RIP, one is typically trying to prevent error in identifying the support of the best noiseless sparse solution of $\tilde{\mathbf{y}} = \boldsymbol{\Upsilon}\boldsymbol{\alpha}$. When constraining observation time and bandwidth, it becomes necessary to use a predetermined grid resolution to satisfy the conditions resulting from the RIP [9], [45]. To contain the modeling error, the only available choice is to abandon the idea of retrieving the ideal $\boldsymbol{\alpha}^\star$ described above, but rather to look for a larger support that sufficiently reduces $\|\boldsymbol{\varepsilon}_{\mathbf{h}}\|_2$. This is what is suggested in [87]. The problem with this approach is that it places too much emphasis on satisfying the conditions resulting from the RIP on the matrix $\boldsymbol{\Upsilon}$ in obtaining the estimate $\hat{\boldsymbol{\alpha}}$ solving (2.13), no matter what the implications are in the estimate $\hat{\mathbf{h}} = \boldsymbol{G}\hat{\boldsymbol{\alpha}}$. However, an alternative, one can still control $\boldsymbol{\varepsilon}_{\mathbf{h}}^\star$ by choosing a finer grid; then, $\|\boldsymbol{\varepsilon}_{\tilde{\mathbf{y}}}\|_2 = \boldsymbol{\mathcal{X}}\|\boldsymbol{\varepsilon}_{\mathbf{h}}^\star\|_2$ will also be small. The residual wiggle-room for spurious solutions satisfying $\|\tilde{\mathbf{y}} - \boldsymbol{\Upsilon}\boldsymbol{\alpha}\|_2^2 < \eta$ should only allow the choice of vectors $\boldsymbol{\alpha}$ such that the system reconstruction error $\|\boldsymbol{G}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)\|^2$ is small. Hence, making errors in the support can be allowed, and the RIP can be dismissed, as long as the resulting error in the estimate of $\mathbf{h}$ is small. A similar observation was made in [6], where the authors were concerned with estimating frequencies of complex sinusoids in noise using a fine uniform quantization of the parameter space and a sparse recovery method. It was demonstrated that by searching for a sparse solution over this grid, CRB level performance on the frequency estimates was obtainable, even with high inter-column correlation,

implying large mutual coherence and restricted isometry constants.

In the next chapter we will discuss in more detail the additional problems encountered when modeling error and noise are present, and propose a metric that can be used to compare input signal designs in terms of their estimation performance.

2.6 Summary

In this chapter, we have proposed a sparse model for MIMO systems that employs a finite grid for the parameter space of our channel (delays, Doppler shifts, angles of arrival and departure), allowing us to use sparse signal recovery methods to estimate the time-varying impulse responses. Using this model, we have examined conditions for sufficiently informative inputs for these models that led to training design guidelines. These guidelines focused on satisfying sufficient conditions such that the channel has a unique sparsest solution over the channel model, implying identifiability for these sparse channels. These conditions are much less restrictive than those associated with the restricted isometry constants or mutual coherence, and allowed for a much finer quantization of the channel parameter space. These conditions, however, do not imply robust designs in the presence of modeling error and noise. In the next chapter we analyze how modeling error and noise will affect the estimate of the channel, and will derive a metric that can be used to compare input signal designs in terms of their corresponding channel estimation performance.

CHAPTER 3

LOCALIZED COHERENCE FOR COMPRESSED CHANNEL SENSING

3.1 Motivation and Related Work

As demonstrated in the previous chapter, at the receiver, the discrete time observations are mapped to a sparse linear combination of columns from an overcomplete matrix. This is similar to the work in [36] and later in [7], [26], [79] for frequency-selective block fading channels, and [9], [57], [87], [60], [45] in the context of doubly-selective SISO channels. In many of these works, the so called *sensing matrix* resulting from the overcomplete basis expansion of the channel combined with the training signal, namely what we denoted as $\boldsymbol{\Upsilon}$ in (2.12) is designed to satisfy certain Restricted Isometry Properties (RIP) or constraints on the mutual coherence (MC) such that the sparse estimation methods (such as Basis Pursuit [24] and Matching Pursuit type methods [64]) guarantee identifiability in the absence of noise and stability in the presence of noise [19], [29]. These sufficient conditions, however, are placing considerable weight in correctly identifying the support of the observations over the sensing matrix and lead to designs that keep the correlation between columns of the sensing matrix at a minimum. Most known families of sensing matrices satisfying the RIP consist of random matrices, such as i.i.d. Gaussian or Bernoulli Matrices, and partial Fourier matrices [72].

The structure of the sensing matrix in the system identification problem cannot be separated from the way the original continuous parameter space of the channel model is discretized. In particular, to ensure low correlation among columns, for a given bandwidth and observation duration, one is implicitly requiring a coarse

discretization of the parameter space. In many practical scenarios, the use of these constraints often leads to systems that cannot be truly modeled as sparse due to modeling error, and lead to a considerable degradation in the channel estimation performance. Inherently, such schemes quantize the parameter space of the channel (delays, Dopplers, angles) and application of the RIP and MC constraints leads to possibly large modeling error in the delay and Doppler parameters. Different from such schemes, the minimal conditions for SI inputs provided in [81], and discussed in the previous chapter, guarantee identifiability of the system when searching for the sparsest system representation, and allow for the use of bases that reduce the overall modeling error by finely quantizing the channel parameter space.

The work in [18] relaxed the RIP and MC conditions when considering the problem of recovering signals from undersampled data, in situations where the signal is not sparse in an incoherent dictionary, but in a truly redundant dictionary. The important aspect of [18] is that it acknowledges that redundancy in the overcomplete channel basis expansion does not necessarily imply problems in estimating it. Our work, like [6, 18], indicates that it is not the high mutual coherence of the channel basis that drives the channel estimation performance, but rather the type of error associated with exchanging specific columns of this basis.

The SI inputs provide guarantees that are not different from the classical definition in [61] and, notably, do not necessarily provide good estimation performance in the presence of noise and/or modeling error. However, it is a worthy effort to search for solutions in this class that lead to robust designs. To rank the robustness of SI signals to noise and modeling errors, we introduce in Sec. 3.4 a new metric, which we refer to as *localized coherence* (LC). We argue that LC is an accurate predictor of estimation performance, and should be preferred over the too restrictive

RIP.

In the following, we will examine how the estimate of the channel support in the quantized parameter space affects the channel estimate. Based on this analysis we derive the metric of *localized coherence*, that can be used to compare different input designs, for a fixed quantization of the parameter space, in order to determine those that lead to better channel estimation performance. To derive our metric, we will use as an example the case of the doubly-selective SISO channel, but note that the analysis is easily extended to the more general models considered in the previous chapter.

In Sec. 3.5, we will apply the metric of localized coherence to the SI input designs discussed in Sec. 2.4, suggesting slight modifications to these designs in order to improve their performance in the presence of noise and modeling error.

3.2 System Model

In the following, we will consider the special case of the MISO system, encountered in the previous chapter. Recall that the observations consisted of the linear combination of Q signals, each corresponding to a particular propagation path described by the path vector $\boldsymbol{\kappa}_q = (\mu_q, f_q, \tau_q)^T$ and attenuation a_q:

$$\mathbf{y} = \sum_q \mathbf{m}(\boldsymbol{\kappa}_q) a_q + \mathbf{w} = \mathbf{M}(\boldsymbol{\kappa})\mathbf{a} + \mathbf{w} \tag{3.1}$$

where $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2 \mathbf{I})$, $\boldsymbol{\kappa} = [\boldsymbol{\kappa}_1^T, ..., \boldsymbol{\kappa}_Q^T]^T$ and $\mathbf{a} = [a_1, ..., a_Q]^T$. For the class of linearly modulated inputs $s_v(t) = \sum_{n=0}^{N-1} p(t - nT_s)$, and where we had assumed the channel was well approximated by a response with memory $L-1$, we had the

following equivalent model:

$$\mathbf{y} = \boldsymbol{\mathcal{X}}\mathbf{h} + \mathbf{w} = \boldsymbol{\mathcal{X}}\mathbf{G}(\boldsymbol{\kappa})\mathbf{a} + \mathbf{w} \tag{3.2}$$

where the matrix $\boldsymbol{\mathcal{X}}$ containing the data symbols, the vector of channel coefficients $\mathbf{h}$, and the basis $\mathbf{G}(\boldsymbol{\kappa})$ of the channel space:

$$\begin{aligned} \{\boldsymbol{\mathcal{X}}\}_{k,l+cL+vLK} &= x_v[k-l]\delta[c-k] \\ \mathbf{h} &= \sum_{q=1}^{Q}(\mathbf{e}(\mu_q)\otimes\mathbf{e}(f_q)\otimes\mathbf{p}(\tau_q))a_q \\ \{\mathbf{G}(\boldsymbol{\kappa})\}_q &= \mathbf{e}(\mu_q)\otimes\mathbf{e}(f_q)\otimes\mathbf{p}(\tau_q) \end{aligned}$$

Special cases of this model included the SISO cases considered in Sec. 2.4.1 and the MISO cases considered in Sec. 2.4.2. We had assumed the channel estimation was performed with the assistance of a training signals. At the receiver, the K observations were projected onto an M-dimensional subspace containing the training, via an $M \times K$ matrix $\mathbf{P}$:

$$\tilde{\mathbf{y}} = \mathbf{P}\mathbf{y} = \mathbf{P}\boldsymbol{\mathcal{X}}\mathbf{G}(\boldsymbol{\kappa})\mathbf{a} + \tilde{\mathbf{w}}. \tag{3.3}$$

In all cases, the matrix $\mathbf{P}\mathbf{P}^H = \mathbf{I}$, such that $\tilde{\mathbf{w}} \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2\mathbf{I})$. As an example, in the time-selective SISO system discussed in Sec. 2.4.1, the matrix $\mathbf{P}$ corresponds to a selection matrix selecting only the M samples of $\mathbf{y}$ containing the pilot symbol. Similarly, in an OFDM system, the matrix $\mathbf{P}$ is the $M \times K$ DFT matrix where K is the number of subcarriers and M the number of pilots used.

We will assume that each parameter vector $\boldsymbol{\kappa}_q$ lies in a bounded region of $\mathbb{R}^3$ where we have placed a finite grid of $R = R_f R_\tau R_\mu$ points $(\mu_{r_\mu}, f_{r_f}, \tau_{r_\tau})$, each point being identified by an index $\mathbf{r} = (r_\mu, r_f, r_\tau)$. Assuming a fine quantization of this

parameter space, and applying Prop. 2.2.1

$$\tilde{\mathbf{y}} = \mathbf{P}\mathbf{M}\boldsymbol{\alpha}^\star + \tilde{\mathbf{w}} + \boldsymbol{\varepsilon}_{\tilde{\mathbf{y}}} = \underbrace{\mathbf{P}\boldsymbol{\mathcal{X}}\boldsymbol{G}}_{\boldsymbol{\Upsilon}}\boldsymbol{\alpha}^\star + \tilde{\mathbf{w}} + \boldsymbol{\varepsilon}_{\tilde{\mathbf{y}}} \tag{3.4}$$

$$\mathbf{h} = \boldsymbol{G}\boldsymbol{\alpha}^\star + \boldsymbol{\varepsilon}_{\mathbf{h}} \tag{3.5}$$

where $\boldsymbol{\varepsilon}_{\tilde{\mathbf{y}}}$ and $\boldsymbol{\varepsilon}_{\mathbf{h}}$ represent the modeling errors resulting from approximating the parameter vectors $\boldsymbol{\kappa}_q$ on the grid. If the parameters $\boldsymbol{\kappa}_q$ fall on the aforementioned grid, $\boldsymbol{\varepsilon}_{\tilde{\mathbf{y}}} = \boldsymbol{\varepsilon}_{\mathbf{h}} = \mathbf{0}$. Alternatively, a possible approximation is obtained considering the Taylor expansion of the model with respect to $\boldsymbol{\kappa}_q$ around the closest grid point, which we assume to exist with finite coefficients; hence, we know there exist vectors $\boldsymbol{\alpha}^\star$ with only Q nonzero elements that result in vanishing modeling error as the grid is made finer. In deriving our analytical results, we assume that each of the parameters $\boldsymbol{\kappa}_q$ has a corresponding point in the grid such that $\|\boldsymbol{\alpha}^\star\|_0 = Q$ with $\boldsymbol{\varepsilon}_{\tilde{\mathbf{y}}} = \boldsymbol{\varepsilon}_{\mathbf{h}} = \mathbf{0}$, unless otherwise stated. We relax this assumption in our numerical analysis. In the following we assume that the system is sparse i.e. that $Q \leq R/2$ elements of $\boldsymbol{\alpha}^\star$ are nonzero. The support of the vector $\boldsymbol{\alpha}^\star$ over $\boldsymbol{\Upsilon}$ (as well as $\boldsymbol{G}$) is denoted as $\mathcal{Q} = (q_1, q_2, ..., q_Q)$. The Q length vector $\boldsymbol{\alpha}^\star_{\mathcal{Q}}$ contains these nonzero components while $\mathbf{u}_{\mathcal{Q}}$ and $\boldsymbol{G}_{\mathcal{Q}}$ are the matrices consisting only of those Q columns.

In the next section, assuming the use of a SI input for this system, we analyze how the channel estimation is affected by the choice of the input signal, when using a sparse recovery algorithm similar to (see [29]):

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \|\boldsymbol{\alpha}\|_0 \text{ subj. to } \|\tilde{\mathbf{y}} - \boldsymbol{\Upsilon}\boldsymbol{\alpha}\|^2 \leq \eta \tag{3.6}$$

where η is, with high probability, an upperbound on the noise level $\|\tilde{\mathbf{w}}\|^2$. Since $\mathbb{E}\{\|\tilde{\mathbf{w}}\|_2^2\} = M\sigma_w^2$, an appropriate choice of η should scale with the noise variance, such as[1] $\eta = C_0 M\sigma_w^2$.

[1]In most sparse algorithms, the choice of C_0 is typically determined via trial and error. We note that choosing C_0 too large or small may significantly deteriorate the estimation performance of the algorithm.

We note that (3.6) is combinatorial in nature and many authors have studied less complex alternatives to this algorithm including greedy approaches such as matching pursuit (MP) and orthogonal matching pursuit (OMP) [64,69], and the l_1 relaxation of (3.6) (obtained by replacing $\|(\cdot)\|_0$ with $\|(\cdot)\|_1$) termed Basis Pursuit or l_1 synthesis [18, 24]. In [18], the authors introduce a criterion and bound on the estimation error of $\mathbf{h}$ using the optimization $\hat{\mathbf{h}} = \arg\min_{\tilde{\mathbf{h}}} \|\boldsymbol{G}^H\tilde{\mathbf{h}}\|_1$ s.t. $\|\tilde{\mathbf{y}} - \mathbf{P}\boldsymbol{\mathcal{X}}\tilde{\mathbf{h}}\|_2^2 \leq \eta$, called l_1 analysis, mentioning in Sec. 4.1 that the l_1 synthesis and analysis problems generally provide very different results. In the parameterizations we consider, the matrix $\boldsymbol{G}^H\mathbf{h}$ is not necessarily sparse. Since (3.6) is based on the assumption of having few propagation paths, with no further indirect mathematical constraints on the model, we will focus on its performance in the following.

3.3 Sparse Recovery Algorithm Error Analysis

In the following we denote $\mathcal{R} = \{1, ..., R\}$ as the set of all column indices of $\boldsymbol{\Upsilon}$ and $2^{\mathcal{R}}$ as its power set. The sparse recovery algorithm we analyze is based on (3.6). We assume that every subset of M columns of the $M \times R$ matrix $\boldsymbol{\Upsilon}$ is linearly independent, which via Def. 2.3.2 is equivalent to saying that the input is sufficiently informative for channels with $\|\boldsymbol{\alpha}^\star\|_0 \leq M/2$. For a deterministic $\boldsymbol{\alpha}$ the assumption that the input is SI implies that, for each set $\mathcal{S}$ with $|\mathcal{S}| \leq N$, the estimate $\hat{\boldsymbol{\alpha}}_{\mathcal{S}} = (\boldsymbol{\Upsilon}_{\mathcal{S}}^H\boldsymbol{\Upsilon}_{\mathcal{S}})^{-1}\boldsymbol{\Upsilon}_{\mathcal{S}}^H\mathbf{y}$ exists since $\boldsymbol{\Upsilon}_{\mathcal{S}}$ is full rank and it minimizes $\|\tilde{\mathbf{y}} - \boldsymbol{\Upsilon}_{\mathcal{S}}\boldsymbol{\alpha}_{\mathcal{S}}\|_2^2$, which in turn means that if the constraint can be met with set $\mathcal{S}$, then $\hat{\boldsymbol{\alpha}}_{\mathcal{S}}$ must be a valid solution. We note that this would be the maximum likelihood solution in AWGN if the support $\mathcal{Q}$ were known apriori, and $\hat{\boldsymbol{\alpha}}_{\mathcal{Q}}$ is sometimes referred to as the Oracle estimate. Replacing $\hat{\boldsymbol{\alpha}}_{\mathcal{S}}$ in the constraint defined in (3.6), and adding the constraint that the solution must satisfy $|\mathcal{S}| \geq 1$,

we get:

$$\tilde{\mathcal{S}} = \underset{\mathcal{S}' \in 2^{\mathcal{R}}}{\operatorname{argmin}} |\mathcal{S}'| \text{ subj. to } \|\boldsymbol{\Pi}_{\mathcal{S}'}^{\perp} \tilde{\mathbf{y}}\|_2^2 \leq \eta, |\mathcal{S}'| \geq 1 \tag{3.7}$$

where the matrix $\boldsymbol{\Pi}_{\mathcal{S}}^{\perp} \triangleq \mathbf{I} - \boldsymbol{\Upsilon}_{\mathcal{S}} (\boldsymbol{\Upsilon}_{\mathcal{S}}^H \boldsymbol{\Upsilon}_{\mathcal{S}})^{-1} \boldsymbol{\Upsilon}_{\mathcal{S}}^H$ is the orthogonal complement to the column-space of $\boldsymbol{\Upsilon}_{\mathcal{S}}$. Due to the assumption that every subset of M columns of $\boldsymbol{\Upsilon}$ is linearly independent, the cardinality of the solution $\tilde{\mathcal{S}}$ will always be such that $|\tilde{\mathcal{S}}| \leq M$. In order to simplify the analysis, we define the sets:

$$\mathcal{C}_1 = \{\mathcal{S} : \|\boldsymbol{\Pi}_{\mathcal{S}}^{\perp} \tilde{\mathbf{y}}\|_2^2 \leq \eta, |\mathcal{S}| \geq 1\}, \quad \mathcal{C}_2 = \{\mathcal{S} : \mathcal{S} = \arg \min_{\mathcal{S}' \in \mathcal{C}_1} |\mathcal{S}'|\} \tag{3.8}$$

where $\mathcal{C}_1$ is the set of supports satisfying the reconstruction error constraint in (3.7) and $\mathcal{C}_2$ is the set of all solutions to (3.7). There may be multiple solutions $\tilde{\mathcal{S}}$ to (3.7) and there are several ways of resolving these ambiguities; for example, selecting an element from $\mathcal{C}_2$ at random or the one minimizing $\|\boldsymbol{\Pi}_{\mathcal{S}}^{\perp} \tilde{\mathbf{y}}\|_2^2$. In the following, we define $\xi(\cdot) : \mathcal{C}_2 \rightarrow \mathcal{C}_2$ as the function resolving this possible ambiguity, and returning a single solution $\hat{\mathcal{S}} = \xi(\mathcal{C}_2)$. The basic algorithm is summarized in Alg. 1.

In summary, the algorithm starts exploring all $\mathcal{S} : \ |\mathcal{S}| = 1$, to determine if any such $\mathcal{S}$ belongs to $\mathcal{C}_1$, including them in a temporary set Θ. If Θ is non empty, then $\mathcal{C}_2 = \Theta$, and the estimated set $\hat{\mathcal{S}} = \xi(\mathcal{C}_2)$. Otherwise, the algorithm starts exploring all $\mathcal{S} : \ |\mathcal{S}| = 2$, and continues the same process until Θ is non empty and $\mathcal{C}_2$ is found. Though this algorithm is combinatorial in nature, its analysis can provide insight into the performance of similar algorithms attempting to find the sparsest solution, while constraining the reconstruction error $\|\tilde{\mathbf{y}} - \boldsymbol{\Upsilon} \hat{\boldsymbol{\alpha}}\|_2^2$.

Our primary concern is not the error $\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^{\star}\|_2$, but the channel estimation error:

$$\|\hat{\mathbf{h}} - \mathbf{h}\|_2 \leq \|(\mathbf{I} - \boldsymbol{G}_{\hat{\mathcal{S}}} \boldsymbol{\Upsilon}_{\hat{\mathcal{S}}}^{\dagger} \mathbf{P} \boldsymbol{\mathcal{X}}) \boldsymbol{G}_{\mathcal{Q}} \boldsymbol{\alpha}_{\mathcal{Q}}\|_2 + \|\boldsymbol{G}_{\hat{\mathcal{S}}} \boldsymbol{\Upsilon}_{\hat{\mathcal{S}}}^{\dagger} \tilde{\mathbf{w}}\|_2 \tag{3.9}$$

Algorithm 1: Sparse Algorithm

$s = 1$, cont $= 0, \Theta = \{\emptyset\}$

while cont $= 1$ **do**

  for $m = 1$ to $\begin{pmatrix} R \\ s \end{pmatrix}$ **do**

    if $\|\mathbf{\Pi}^{\perp}_{\mathcal{S}^{(m)}}\mathbf{y}\|_2^2 \leq \eta$ **then**

      $\Theta \Leftarrow \{\Theta, \mathcal{S}^{(m)}\}$

    end if

  end for

  if $\Theta \neq \{\emptyset\}$ **then**

    $\mathcal{C}_2 \Leftarrow \Theta$, cont $= 0$

  else

    $s = s + 1$

  end if

end while

$\hat{\mathcal{S}} = \xi(\mathcal{C}_2)$, $\hat{\mathbf{h}} = \boldsymbol{G}_{\hat{\mathcal{S}}}\hat{\boldsymbol{\alpha}}_{\hat{\mathcal{S}}}$

where $\mathbf{\Upsilon}^{\dagger}_{\hat{\mathcal{S}}}$ is the pseudo-inverse of $\mathbf{\Upsilon}_{\hat{\mathcal{S}}}$. The first term is the error in modeling the channel due to the particular estimate of the columns $\hat{\mathcal{S}}$, while the second term is the noise entering the system. For any $\hat{\mathcal{S}} \supseteq \mathcal{Q}$, the first term is zero, however, large values of $\hat{\mathcal{S}}$ generally result in larger values of the second term and for this reason sparse estimation methods have the added benefit of denoising the estimate.

Though the channel MSE is difficult to analyze, due to the nonlinear dependence of $\hat{\mathcal{S}}$ on $\tilde{\mathbf{y}}$, we will use the following bound on the normalized mean Euclidean norm error (NMEE) to gauge how the performance is affected by the choice of the input signal. Defining $C = \|\boldsymbol{\alpha}_{\mathcal{Q}}\|_2\sqrt{M}$, and noting that

$\|\mathbf{A}\boldsymbol{\alpha}\|_2 \leq \sqrt{\lambda_{max}(\mathbf{A}^H\mathbf{A})}\|\boldsymbol{\alpha}\|_2$, we have:

$$\frac{\mathbb{E}\{\|\hat{\mathbf{h}} - \mathbf{h}\|_2\}}{\|\boldsymbol{\alpha}_{\mathcal{Q}}\|_2\sqrt{M}} \leq \frac{\mathbb{E}\{d(\mathcal{Q},\hat{\mathcal{S}})\}}{\sqrt{M}} + \frac{\mathbb{E}\{\|\boldsymbol{G}_{\hat{\mathcal{S}}}\boldsymbol{\Upsilon}^{\dagger}_{\hat{\mathcal{S}}}\tilde{\mathbf{w}}\|_2\}}{C} \tag{3.10}$$

where $d(\mathcal{Q},\hat{\mathcal{S}}) = \sqrt{\lambda_{\max}(\boldsymbol{G}^H_{\mathcal{Q}}\mathbf{U}^H_{\hat{\mathcal{S}}}\mathbf{U}_{\hat{\mathcal{S}}}\boldsymbol{G}_{\mathcal{Q}})}$ with $\mathbf{U}_{\hat{\mathcal{S}}} = \mathbf{I} - \boldsymbol{G}_{\hat{\mathcal{S}}}\boldsymbol{\Upsilon}^{\dagger}_{\hat{\mathcal{S}}}\mathbf{P}\boldsymbol{\mathcal{X}}$, and the expectation is with respect to $\hat{\mathcal{S}}$ and $\tilde{\mathbf{w}}$ and conditioned on $\boldsymbol{\alpha}$. The first term in the bound captures the error in modeling the channel due to errors in the support selection, while the second term is tied to the effects of noise in the channel estimate. Clearly, replacing $\hat{\mathcal{S}}$ with any of its subsets would reduce the contribution of the second term. Thus, a good estimator should strike a balance between increasing the modeling error and decreasing the noise in the estimate by picking a small model size $|\hat{\mathcal{S}}|$. It is important to remark that a good estimate does not necessarily require resolving all of the Q paths and the corresponding columns of $\boldsymbol{G}_{\mathcal{Q}}$, but rather from selecting a $\boldsymbol{G}_{\hat{\mathcal{S}}}$ whose column space well approximates that of $\boldsymbol{G}_{\mathcal{Q}}$. In the following we will analyze the first term[2] in (3.10) to find a criterion that could be used as an alternative to the RIP and/or MC requirements. Specifically:

$$e(\mathcal{Q}) = \mathbb{E}\{d(\mathcal{Q},\mathcal{S})\} = \sum_{m=1}^{M}\sum_{\mathcal{S}:|\mathcal{S}|=m} d(\mathcal{Q},\mathcal{S})p_{\hat{\mathcal{S}}}(\mathcal{S}) \tag{3.11}$$

In defining the probability law $p_{\hat{\mathcal{S}}}(\mathcal{S}) \triangleq \Pr(\mathcal{S} = \xi(\mathcal{C}_2))$, the random experiment consists of performing Alg. 1 on noisy observations such as (3.4) and the sample space Ω consists of the power set $\Omega = 2^{\mathcal{R}}$. Determining the probability of the random outcome $\hat{\mathcal{S}}$ in closed form is difficult. We note that $\Pr(\mathcal{S} = \xi(\mathcal{C}_2)) = \Pr((\mathcal{S} = \xi(\mathcal{C}_2)) \cap (\mathcal{S} \in \mathcal{C}_2)) \leq \Pr(\mathcal{S} \in \mathcal{C}_2)$. In a similar fashion, as $\mathcal{C}_2 \subseteq \mathcal{C}_1$, $\Pr(\mathcal{S} \in \mathcal{C}_2) = \Pr((\mathcal{S} \in \mathcal{C}_2) \cap (\mathcal{S} \in \mathcal{C}_1)) = \Pr(\mathcal{S} \in \mathcal{C}_2|\mathcal{S} \in \mathcal{C}_1)\Pr(\mathcal{S} \in \mathcal{C}_1)$:

$$e(\mathcal{Q}) \leq \sum_{m=1}^{M}\sum_{\mathcal{S}:|\mathcal{S}|=m} d(\mathcal{Q},\mathcal{S})\Pr(\mathcal{S} \in \mathcal{C}_2|\mathcal{S} \in \mathcal{C}_1)\Pr(\mathcal{S} \in \mathcal{C}_1) \tag{3.12}$$

[2]The analysis of the second term is complicated due to the fact that $\hat{\mathcal{S}}$ and $\tilde{\mathbf{w}}$ are dependent. One could bound the second term by a constant times $M\sigma_w^2$, but this would not provide any additional insight on the error trend and we will not consider the second term for this reason.

We observe that:

$$\Pr(\mathcal{S} \in \mathcal{C}_2 | \mathcal{S} \in \mathcal{C}_1) = 1 - \Pr(\mathcal{S} \notin \mathcal{C}_2 | \mathcal{S} \in \mathcal{C}_1) = 1 - \Pr(\min_{\mathcal{S}' \in \mathcal{C}_1} |\mathcal{S}'| < |\mathcal{S}|) \quad (3.13)$$

Thus $\Pr(\mathcal{S} \in \mathcal{C}_2 | \mathcal{S} \in \mathcal{C}_1)$ is monotonically decreasing in $|\mathcal{S}|$ and is unity for all $|\mathcal{S}| = 1$, as we have restricted $|\hat{\mathcal{S}}| \geq 1$. Intuitively, we expect that for larger values of η in (3.7), this probability will decrease faster in $|\mathcal{S}|$, while for very small values of η it remains large for all $|\mathcal{S}| \leq M$.

Assuming $\tilde{\mathbf{y}} \sim \mathcal{CN}(\boldsymbol{\mu}_{\tilde{\mathbf{y}}}, \boldsymbol{\Sigma})$, the term $\Pr(\mathcal{S} \in \mathcal{C}_1)$ is easily upperbounded using the Chernoff bound:

$$\Pr(\mathcal{S} \in \mathcal{C}_1) = \Pr(\|\boldsymbol{\Pi}_{\mathcal{S}}^{\perp} \tilde{\mathbf{y}}\|^2 \leq \eta) \leq \frac{e^{\eta}}{\det(\boldsymbol{\Pi}_{\mathcal{S}}^{\perp} \boldsymbol{\Sigma} + \mathbf{I})} e^{-\boldsymbol{\mu}_{\tilde{\mathbf{y}}}^H (\boldsymbol{\Pi}_{\mathcal{S}}^{\perp} (\mathbf{I} + \boldsymbol{\Sigma} \boldsymbol{\Pi}_{\mathcal{S}}^{\perp})^{-1}) \boldsymbol{\mu}_{\tilde{\mathbf{y}}}} \quad (3.14)$$

In contrast to (3.13), we note that (3.14) is increasing in η. We use (3.14) under two assumptions: first that $\boldsymbol{\alpha}_{\mathcal{Q}}$ and $\mathcal{Q}$ are deterministic, where $\boldsymbol{\mu}_{\tilde{\mathbf{y}}} = \boldsymbol{\Upsilon}_{\mathcal{Q}} \boldsymbol{\alpha}_{\mathcal{Q}}$ and $\boldsymbol{\Sigma} = \sigma_w^2 \mathbf{I}$, and second, that $\mathcal{Q}$ is deterministic but $\boldsymbol{\alpha}_{\mathcal{Q}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}\sigma_a^2)$ and independent of $\tilde{\mathbf{w}}$ where $\boldsymbol{\mu}_{\tilde{\mathbf{y}}} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \boldsymbol{\Upsilon}_{\mathcal{Q}} \boldsymbol{\Upsilon}_{\mathcal{Q}}^H \sigma_a^2 + \sigma_w^2 \mathbf{I}$. We will relate these two probability measures later to extract our *localized coherence* metric. Denoting $\pi^{(d)}$ and $\pi^{(r)}$ as (3.14) in these different cases, respectively, we have:

$$\pi^{(d)} \leq e^{\eta} e^{-\boldsymbol{\alpha}_{\mathcal{Q}}^H \boldsymbol{\Upsilon}_{\mathcal{Q}}^H (\boldsymbol{\Pi}_{\mathcal{S}}^{\perp} (\mathbf{I} + \sigma_w^2 \boldsymbol{\Pi}_{\mathcal{S}}^{\perp})^{-1}) \boldsymbol{\Upsilon}_{\mathcal{Q}} \boldsymbol{\alpha}_{\mathcal{Q}}} (1 + \sigma_w^2)^{|\mathcal{S}| - N} \quad (3.15)$$

$$\pi^{(r)} \leq e^{\eta} \det(\boldsymbol{\Pi}_{\mathcal{S}}^{\perp} (\boldsymbol{\Upsilon}_{\mathcal{Q}} \boldsymbol{\Upsilon}_{\mathcal{Q}}^H \sigma_a^2 + \mathbf{I}\sigma_w^2) + \mathbf{I})^{-1} \quad (3.16)$$

For small values of σ_w^2, the second exponential term in (3.15) can be approximated as $e^{-\|\boldsymbol{\Pi}_{\mathcal{S}}^{\perp} \boldsymbol{\Upsilon}_{\mathcal{Q}} \boldsymbol{\alpha}_{\mathcal{Q}}\|_2^2}$ such that (3.15) is largest when the column space of $\boldsymbol{\Upsilon}_{\mathcal{S}}$ spans that of $\boldsymbol{\Upsilon}_{\mathcal{Q}}$. Likewise, in (3.16), where $\boldsymbol{\alpha}_{\mathcal{Q}}$ is random, for small σ_w^2 we observe that the largest values of (3.16) similarly correspond to those cases where the column space of $\boldsymbol{\Upsilon}_{\mathcal{S}}$ spans that of $\boldsymbol{\Upsilon}_{\mathcal{Q}}$.

Combining the expressions (3.13) and (3.14), we have the following upperbounds on $e(\mathcal{Q})$. For the case of deterministic $\boldsymbol{\alpha}_{\mathcal{Q}}$ we have $e(\mathcal{Q}) \leq \sum_{m=1}^{M} \phi^{(d)}(m)$,

where

$$\phi^{(d)}(m) = \begin{cases} \sum\limits_{\mathcal{S}:|\mathcal{S}|=m} \frac{e^{\eta d(\mathcal{Q},\mathcal{S})}(1+\sigma_w^2)^{m-M}}{e^{\boldsymbol{\alpha}_\mathcal{Q}^H \boldsymbol{\Upsilon}_\mathcal{Q}^H (\boldsymbol{\Pi}_\mathcal{S}^\perp (\mathbf{I}+\sigma_w^2 \boldsymbol{\Pi}_\mathcal{S}^\perp)^{-1}) \boldsymbol{\Upsilon}_\mathcal{Q} \boldsymbol{\alpha}_\mathcal{Q}}}, & m=1 \\ \sum\limits_{\mathcal{S}:|\mathcal{S}|=m} \frac{e^{\eta d(\mathcal{Q},\mathcal{S})}(1+\sigma_w^2)^{m-M}(1-\Pr(\min\limits_{\mathcal{S}\in\mathcal{C}_1}|\mathcal{S}'|<m))}{e^{\boldsymbol{\alpha}_\mathcal{Q}^H \boldsymbol{\Upsilon}_\mathcal{Q}^H (\boldsymbol{\Pi}_\mathcal{S}^\perp (\mathbf{I}+\sigma_w^2 \boldsymbol{\Pi}_\mathcal{S}^\perp)^{-1}) \boldsymbol{\Upsilon}_\mathcal{Q} \boldsymbol{\alpha}_\mathcal{Q}}}, & m>1 \end{cases} \tag{3.17}$$

In the case of random $\boldsymbol{\alpha}_\mathcal{Q} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}\sigma_a^2)$ we have $e(\mathcal{Q}) \leq \sum_{m=1}^M \phi^{(r)}(m)$, where

$$\phi^{(r)}(m) = \begin{cases} \sum\limits_{\mathcal{S}:|\mathcal{S}|=m} \frac{e^{\eta d(\mathcal{Q},\mathcal{S})}}{\det(\boldsymbol{\Pi}_\mathcal{S}^\perp (\boldsymbol{\Upsilon}_\mathcal{Q} \boldsymbol{\Upsilon}_\mathcal{Q}^H \sigma_a^2 + \mathbf{I}\sigma_w^2)+\mathbf{I})}, & m=1 \\ \sum\limits_{\mathcal{S}:|\mathcal{S}|=m} \frac{e^{\eta d(\mathcal{Q},\mathcal{S})}(1-\Pr(\min\limits_{\mathcal{S}\in\mathcal{C}_1}|\mathcal{S}'|<m))}{\det(\boldsymbol{\Pi}_\mathcal{S}^\perp (\boldsymbol{\Upsilon}_\mathcal{Q} \boldsymbol{\Upsilon}_\mathcal{Q}^H \sigma_a^2 + \mathbf{I}\sigma_w^2)+\mathbf{I})}, & m>1 \end{cases} \tag{3.18}$$

For appropriately chosen η, the highest probability sets $\mathcal{S}$ are those where the columns $\boldsymbol{\Upsilon}_\mathcal{S}$ closely span the column space of $\boldsymbol{\Upsilon}_\mathcal{Q}$ while maintaining minimum cardinality. We can infer that an appropriate design of the input signal should try to ensure $d(\mathcal{Q}, \mathcal{S})$ is small for these high probability sets $\mathcal{S}$. As it is intractable to analyze these bounds for all m, we examine the case of $m = 1$ assuming $|\mathcal{Q}| = 1$ and for negligible noise $\sigma_w^2 \approx 0$. For $|\mathcal{S}| = 1$ and $|\mathcal{Q}| = 1$, we have $\mathcal{S} = \{s\}$ and $\mathcal{Q} = \{q\}$ and we define $\mathbf{u}_q \triangleq \boldsymbol{\Upsilon}_{\{q\}}$, $\mathbf{u}_s \triangleq \boldsymbol{\Upsilon}_{\{s\}}$ and $\alpha_q = \boldsymbol{\alpha}_{\{q\}}$, and have the following approximations

$$\phi^{(d)}(1) \approx \sum_{\substack{\{s\}\in\mathcal{R} \\ \{s\}\neq\{q\}}} d(\{q\},\{s\}) e^\eta \underbrace{e^{-\|\alpha_q\|_2^2 \left(\|\mathbf{u}_q\|_2^2 - \frac{|\mathbf{u}_q^H \mathbf{u}_s|^2}{\|\mathbf{u}_s\|_2^2} \right)}}_{p_\phi^{(d)}(\mathbf{u}_q, \mathbf{u}_s)} \tag{3.19}$$

$$\phi^{(r)}(1) \approx \sum_{\substack{\{s\}\in\mathcal{R} \\ \{s\}\neq\{q\}}} d(\{q\},\{s\}) e^\eta \underbrace{\left(1 + \sigma_a^2 (\|\mathbf{u}_q\|^2 - \frac{|\mathbf{u}_q^H \mathbf{u}_s|^2}{\|\mathbf{u}_s\|_2^2}) \right)^{-1}}_{p_\phi^{(r)}(\mathbf{u}_q, \mathbf{u}_s)} \tag{3.20}$$

When each column of $\boldsymbol{\Upsilon}$ has approximately $\|\mathbf{u}_s\|_2^2 \approx c$, the $\mathbf{u}_s$ with largest $p_\phi^{(d)}(\mathbf{u}_q, \mathbf{u}_s)$ and $p_\phi^{(r)}(\mathbf{u}_q, \mathbf{u}_s)$ are those that are highly correlated with $\mathbf{u}_q$. For example if $|\mathbf{u}_q^H \mathbf{u}_s|^2 \approx c^2$, the terms $p_\phi^{(d)}(\mathbf{u}_q, \mathbf{u}_s) \approx 1$ and $p_\phi^{(r)}(\mathbf{u}_q, \mathbf{u}_s) \approx 1$. On the other hand when $\mathbf{u}_q$ and $\mathbf{u}_s$ are approximately orthogonal, $p_\phi^{(d)}(\mathbf{u}_q, \mathbf{u}_s) = e^{-c\|\alpha_q\|_2^2}$ and $p_\phi^{(r)}(\mathbf{u}_q, \mathbf{u}_s) \approx (1 + \sigma_a^2 c)^{-1}$. For $\|\alpha_q\|_2^2 c \ll 1$, we note $e^{-c\|\alpha_q\|^2} \approx (1 + \|\alpha_q\|_2^2 c)^{-1}$.

It is worth noting that this analysis can be applied to other problems, such as the parameter estimation problem, where instead of the channel impulse response, we may be interested in estimating the parameters $\boldsymbol{\kappa}_q$. Since each column of $\boldsymbol{\Upsilon}$ corresponds to a particular set of parameters $\boldsymbol{\kappa}_q$, we can replace $d(\mathcal{Q}, \mathcal{S})$ with $\|\mathbf{U}(\boldsymbol{\kappa}_{\mathcal{Q}} - \boldsymbol{\kappa}_{\mathcal{S}})\|^2$, i.e. the weighted error between the parameters, where $\mathbf{U}$ is the weighting matrix. Even in the channel estimation problem, it can be demonstrated via a Taylor series expansion that small values of $\|\boldsymbol{\kappa}_{\mathcal{Q}} - \boldsymbol{\kappa}_{\mathcal{S}}\|^2$ imply small values of $d(\mathcal{Q}, \mathcal{S})$, suggesting that accurate estimation of the parameters $\boldsymbol{\kappa}_q$ implies accurate estimation of the channel impulse response, as expected.

3.4 Localized Coherence

In Sec. 3.2, we assumed that the quantization of the parameter space of $\boldsymbol{\kappa}_q$ lead to a sparse system, which we later used as if it were the exact representation of the observations. The dilemma encountered is whether or not to refine the quantization when the parameters $\boldsymbol{\kappa}_q$ do not truly lie on the grid. Though a fine grid generally results in larger mutual coherence, it is easy to numerically verify that a fine grid is preferable to a coarse grid, even when using suboptimal approximations to (3.7), such as OMP or Basis Pursuit, as demonstrated in Fig. 3.1. This has also been observed in [6], [80] and indicates that RIP and MC constraints should be complemented with other tools. The issue still remains open as to what constitutes a good input signal, once a fine grid has been chosen and the RIP and MC constraints are no longer satisfied. We use the previous analysis of the NMEE to supply a possible guideline.

From the preceding analysis we observed that, in high SNR situations and

assuming an appropriate choice of η, the most likely errors made in trying to find a sparse solution correspond to the selection of sets of columns $\mathcal{S}$ of $\mathbf{U}$ spanning a space very similar to that spanned by the set of columns $\mathcal{Q}$, and that have similar or smaller support. To reduce the corresponding error, the input design should be such that if $\boldsymbol{\Upsilon}_{\mathcal{S}}$ closely spans that of $\boldsymbol{\Upsilon}_{\mathcal{Q}}$, then the corresponding values of $d(\mathcal{Q},\mathcal{S})$ are small.

We observe that in the cases where σ_w^2 is small, the terms $\det(\boldsymbol{\Pi}_{\mathcal{S}}^{\perp}(\boldsymbol{\Upsilon}_{\mathcal{Q}}\boldsymbol{\Upsilon}_{\mathcal{Q}}^H\sigma_a^2 + \mathbf{I})^{-1}$ and $e^{-\|\boldsymbol{\Pi}_{\mathcal{S}}^{\perp}\boldsymbol{\Upsilon}_{\mathcal{Q}}\boldsymbol{\alpha}_{\mathcal{Q}}\|_2^2}$ are exponentially decreasing in $|\mathcal{Q}|$. This combined with the fact that $\Pr(\mathcal{S}\in\mathcal{C}_2|\mathcal{S}\in\mathcal{C}_1)$ is decreasing in $|\mathcal{S}|$ seems to suggest that analyzing only $p_{\phi}^{(d)}(1)$ and $p_{\phi}^{(r)}(1)$ for $|\mathcal{Q}|=1$ is sufficient to compare the performance for different input designs. In this simplest case with $|\mathcal{Q}|=1$, we observed in (3.19) and (3.20) that the most likely candidates $\mathcal{S}$ (with $|\mathcal{S}|=1$) to be returned by the algorithm are those with large correlation $\|\mathbf{u}_q^H\mathbf{u}_s\|^2$ and we desire for the largest values of inter-column correlation, $|\mathbf{u}_q^H\mathbf{u}_s|^2$, to correspond to small values of $d(\{q\},\{s\})$. We propose a simple measure (called localized coherence) of this property based on (3.20) and (3.19):

$$\psi = \sum_{q=1}^{R}\sum_{s\neq q}\frac{d(\{q\},\{s\})e^{\eta}}{RM\left(1+a(\|\mathbf{u}_q\|_2^2-\frac{|\mathbf{u}_q^H\mathbf{u}_s|^2}{\|\mathbf{u}_s\|^2})\right)} \tag{3.21}$$

where $a=\|\alpha_q\|^2$ for deterministic $\boldsymbol{\alpha}_{\mathcal{Q}}$ ($a=\sigma_a^2$ if random) and $\mathbf{u}_q$ is the qth column of $\boldsymbol{\Upsilon}=\boldsymbol{\mathcal{X}}\boldsymbol{G}$. We want the error to be small for all possible values of q, and is why we sum over q. Localized coherence is a simple measure to aid in designing the input signal (in $\boldsymbol{\mathcal{X}}$) to attain good performance in the presence of noise. Though this metric does not consider the case of more than a single column, it is obviously a starting point for designing $\boldsymbol{\mathcal{X}}$.

Even if most sparse recovery algorithms try to approximate the solution of (3.6), we have observed that this metric is applicable to predict the ranking of

different input design in terms of the resulting channel estimation performance. For example greedy algorithms such as MP and OMP select at each iteration of the algorithm, the column of $\boldsymbol{\Upsilon}$ that is most correlated with the residual signal. Other methods that relax (3.6) by replacing the l_0 norm with the l_1 norm and possibly modifying the reconstruction error constraint, such as Basis Pursuit and the Dantzig Selector [24], [22], appear to similarly select a set of columns $\boldsymbol{\Upsilon}_{\mathcal{S}}$ which span a space very similar to the columns $\boldsymbol{\Upsilon}_{\mathcal{Q}}$.

3.4.1 Numerical Examples

In the following examples, we will use OMP [69] to estimate the coefficients $\boldsymbol{\alpha}^{\star}$ and support $\mathcal{Q}$. We stop the OMP algorithm when the reconstruction error $\|\tilde{\mathbf{y}} - \boldsymbol{\Upsilon}\boldsymbol{\alpha}\|_2^2$ is less than $\eta = M\sigma_w^2$. In all experiments we define the received signal-to-noise ratio (SNR) as $\mathbb{E}\{\|\boldsymbol{\mathcal{X}}\mathbf{h}\|^2\}/\mathbb{E}\{\|\mathbf{w}\|\}^2$. In these examples we will consider the doubly-selective SISO system, assuming a linear modulated signal $s(t) = \sum_{n=0}^{N-1} x[n]p(t - nT_s)$. We will use all available degrees of freedom in the training signal such that $M = N$, and the projection matrix $\mathbf{P} = \mathbf{I}$.

In the first example we compare the channel estimation performance for two different scenarios of the SISO doubly-selective channel. The first scenario assumes that the parameters of the channel $\boldsymbol{\kappa}_q$ lie on the grid of the quantized parameter space while the second scenario assumes the parameters $\boldsymbol{\kappa}_q$ do not lie on the grid points but in a bounded region of $\mathbb{R}^2$. We assume that the baseband equivalent discrete time channel impulse response has $L = 6$ taps. We observe the signal over $K = 17$ observations where the $M = N = 17$ length input sequence consists of two chirp sequences, $x[n] = e^{-j2\pi n\beta/(KT_s^2)}$ for $n = 0, ..8$ and $x[n] = e^{j2\pi n\beta/(KT_s)^2}$ for $n = 9, ..., 16$ and uses a cyclic prefix of length $N_{cp} = L$ (see [81]). The parameter

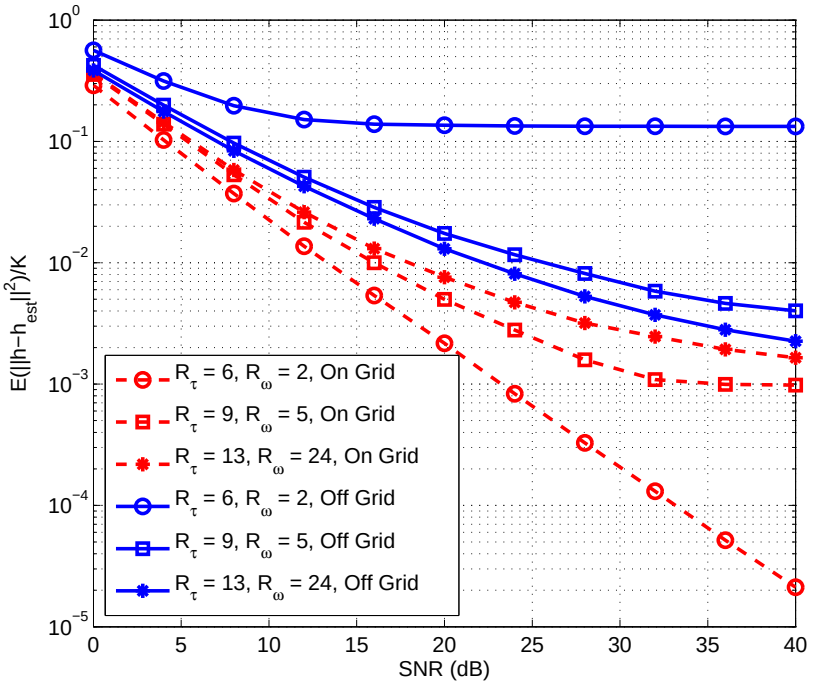

Figure 3.1: Channel MSE for various Grid choices

$\beta = 1.9$ and the pulse $p(t) = \text{sinc}(\pi t/T_s)$.

The channel parameters $\boldsymbol{\kappa}_q = (\tau_q, f_q) \in (0, (L-1)T_s) \times (-.025, .025)$. We construct a uniform grid for the parameter space such that for a given R_τ and R_ω the spacing between delays $\epsilon_\tau = (L-1)T_s/(R_\tau - 1)$ while the spacing between frequencies $\epsilon_\omega = .05/(R_\omega - 1)$. In both scenarios, the path attenuations $\mathbf{a}_\mathcal{Q} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}Q^{-}1)$ with $Q = 2$ paths. In the first scenario, since we assume that the parameters $\boldsymbol{\kappa}_q$ are generated on the grid, we randomly choose Q points in the grid. In the second scenario, the frequencies $f_q \overset{iid}{\sim} \mathcal{U}(-0.025, 0.025)$ while the delays $\tau_q \overset{iid}{\sim} \mathcal{U}(0, (L-1)T_s)$. The system is generated according to (3.2).

In Fig. 3.1 we plot the channel MSE, $\mathbb{E}\{\|\hat{\mathbf{h}} - \mathbf{h}\|^2/K\}$. As expected, when the parameters $\boldsymbol{\kappa}_q$ are generated on the grid and the grid becomes coarser, the columns

of $\boldsymbol{\Upsilon}$ and $\boldsymbol{G}$ are less correlated and thus the estimation performance improves. However, when the parameters are not assumed to lie on the grid, the opposite occurs, and the performance improves as the quantization of the parameter space is made finer. The grid with the best performance in the first scenario corresponds to the worst performance in the second scenario and vice versa. In practice we fall into the second scenario, and is precisely why we consider a fine quantization of the parameter space and propose the use of localized coherence to compare designs of the input sequence for a fixed quantization of the parameter space. Though not shown, when simulating the NMEE, we make the same observations.

We also observe in Fig. 3.1 that as the grid size is increased, the gain in performance appears to saturate while the computational complexity increases. Indeed, we do expect diminishing returns in the second scenario when increasing the grid size. Localized coherence is not quite able to capture this behavior due to the union type bound used. Further, we expect this behavior to be even more prevalent when using sub-optimal methods like OMP compared to (3.7), as done in this experiment.

In the second example, for a fixed grid, we compare the localized coherence ψ (shown in Fig. 3.2) for several different designs of the sequence $x[n]$. The first sequence is the chirp sequence defined in the previous experiment, the second sequence is the Alltop (see [45]) defined as $x[n] = e^{j\pi n^3/N}$ where $M = N = 17$, and the last two sequences are two different binary sequences:

$$\begin{aligned} \text{Binary}(1), \mathbf{s} &= (1,1,1,1,-1,-1,-1,-1,1,1,1,1,-1,-1,-1,-1,-1)^T \\ \text{Binary}(2), \mathbf{s} &= (1,-1,-1,1,1,-1,-1,-1,1,1,1,1,-1,-1,-1,1,1)^T \end{aligned} \quad (3.22)$$

We set the number of observations $K = 17$, and again assume each of the sequences is cyclic prefixed. The same channel parameters are used as in the previous ex-

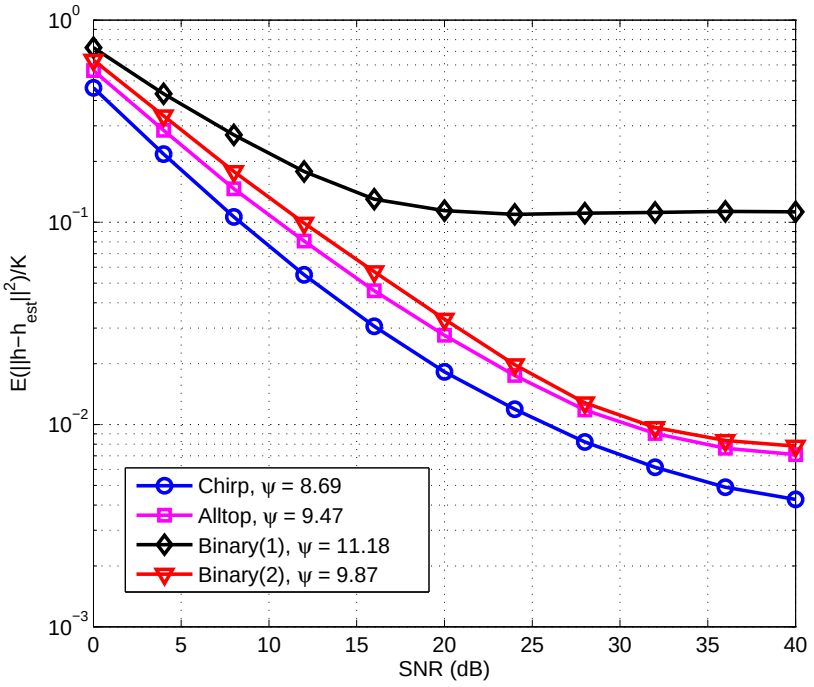

Figure 3.2: Channel MSE for various training sequence designs.

periment, however with $Q = 3$ paths, generating the parameters uniformly at random in the respective intervals. The grid parameters are fixed as $R_\tau = L + 7$ and $R_\omega = 24$, and ϵ_τ and ϵ_ω are defined as in the previous experiment. In Fig. 3.2, we plot the channel MSE, while in Fig. 3.3 we plot the channel NMEE ($\mathbb{E}\{\|\hat{\mathbf{h}} - \mathbf{h}\|_2 \|\mathbf{a}\|_2^{-1}\}/\sqrt{K}$) as well as the first term of (3.10) averaged over $\mathcal{Q}$ and $\mathcal{S}$, (denoted as $\mathbb{E}(d)$ where we have $\boldsymbol{G}_{\mathcal{Q}}$ with $\mathbf{G}(\boldsymbol{\kappa})$ since our parameters are not necessarily on the grid) averaged over 10000 channel realizations.

In Fig.'s 3.2 and 3.3, as expected, the sequences having the best channel MSE, NMEE, and $\mathbb{E}(d)$ performance also have the smallest values of localized coherence ψ. The chirp sequence and the first binary sequences appear to have similar performance, while the second binary sequence and cubic sequence are slightly worse. We also observe that the trend of $\mathbb{E}(d)$ is similar to that of the channel MSE and

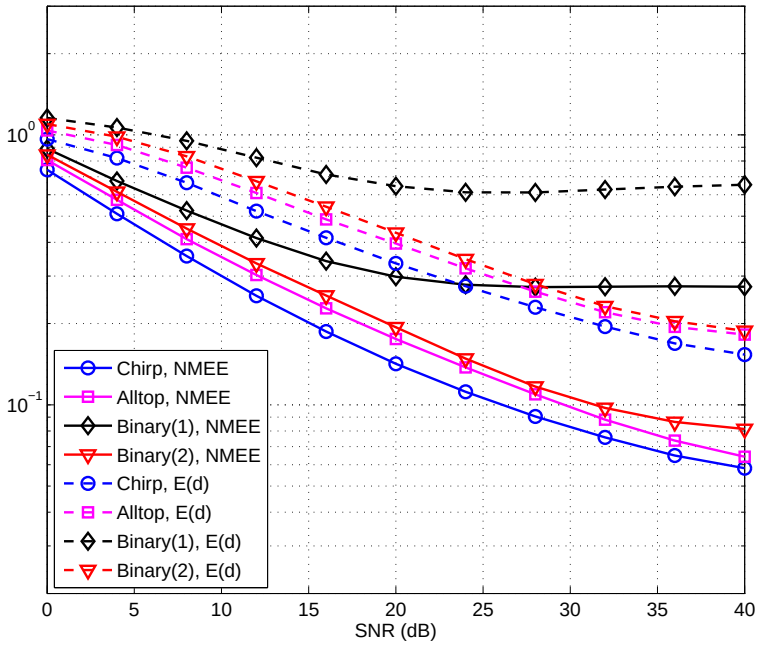

Figure 3.3: Channel NMEE and $\mathbb{E}\{d(\mathcal{Q}, \mathcal{S})\}$ for various training sequence designs.

NMEE, lending further support to the argument that one should try to decrease the contribution from the term $\mathbb{E}(d)$.

Though not shown, we observe similar trends for larger values of Q as well. We have further observed that the MSE performance (as well as NMEE and $\mathbb{E}(d)$) tend to saturate at high values of SNR when using the OMP algorithm. This is to be expected as these input sequence designs generally do not satisfy the conditions for identifiability of the channel (based on the MC) for the OMP algorithm, and thus residual channel estimation error will be present even in the absence of noise.

3.5 Improving Localized Coherence for SI Inputs

In this section, we apply the metric of localized coherence introduced in Sec. 3.4 to the sufficiently informative input designs discussed in Sec. 2.4 leading to modifications in the design of these inputs.

Recall that in the SI designs considered in Sec.'s 2.4.1 and 2.4.2, the matrix $\boldsymbol{\Upsilon}$ always had a special structure, and the observations $\tilde{\mathbf{y}}$ were mapped into linear combinations of columns from $\boldsymbol{\Upsilon}$. Denoting $\mathbf{u_r}$ as the column of $\boldsymbol{\Upsilon}$ associated with the grid point indexed by $\mathbf{r}$ and the path parameters $\boldsymbol{\kappa}_\mathbf{r}$, in these designs each column $\mathbf{u_r}$ is a complex exponential whose frequency is a linear function of the grid point parameter vector $\boldsymbol{\kappa}_\mathbf{r}$. The designs proposed were oblivious to the connection between the correlation of columns of $\boldsymbol{\Upsilon}$ and the MSE performance. In fact, the only constraint imposed was that the input and grid ensured every set of $2S$ columns of $\boldsymbol{\Upsilon}$ were linearly independent. Defining $p(\mathbf{u_r}, \mathbf{u_{r'}}) \triangleq (1 + a(\|\mathbf{u_r}\|_2^2 - \frac{|\mathbf{u_r}^H \mathbf{u_{r'}}|^2}{\|\mathbf{u_{r'}}\|^2}))^{-1}$, we observed that if the columns of $\boldsymbol{\Upsilon}$ have approximately constant norm, than the term $p(\mathbf{u_r}, \mathbf{u_{r'}})$ increases with $|\mathbf{u_r}^H \mathbf{u_{r'}}|^2$. Since $\mathbf{u_r}$ and $\mathbf{u_{r'}}$ are complex sinusoids in Sec.'s 2.4.1 and 2.4.2, $p(\mathbf{u_r}, \mathbf{u_{r'}})$ increases as the two sinusoids get closer in frequency.

For these designs, we can interpret localized coherence in the following manner. Each column $\mathbf{u_r}$ has a neighborhood consisting of those $\mathbf{u_{r'}}$ such that $|\mathbf{u_r}^H \mathbf{u_{r'}}|^2 > C$, where C is a real number specifying the amount of correlation[3] used in defining the neighborhood. On the other hand, since it easily shown that $d(\mathbf{r}, \mathbf{r'})$ is decreasing in $\|\mathbf{W}(\boldsymbol{\kappa}_\mathbf{r} - \boldsymbol{\kappa}_{\mathbf{r'}})\|_2^2$, where $\mathbf{W}$ is an appropriately defined weighting matrix, an indirect way to keep the LC ψ small is to ensure that neighbors always correspond

[3]We use correlation to define the neighborhood, as opposed to Euclidean distance, as our coefficients $\mathbf{a}$ are unknown.

to small values of $\|\mathbf{W}(\boldsymbol{\kappa}_{\mathbf{r}} - \boldsymbol{\kappa}_{\mathbf{r}'})\|_2^2$. This is precisely the idea that we use next to improve upon the SI designs presented in 2.4.

Some of the cases presented in 2.4 are more problematic than others. It is relatively simple to improve the LC for the time and frequency-selective SI designs considered in Sec. 2.4.1, as the proposed pilot signals resulted in matrices $\boldsymbol{\Upsilon}$ where each column was a complex sinusoid whose frequency was proportional to the parameter (a model very similar to the model in [6]). The larger the spacing among frequencies, the smaller $p(\mathbf{u}_{\mathbf{r}}, \mathbf{u}_{\mathbf{r}'})$ is and therefore the better the neighborhoods. For the single parameter cases we considered, assuming uniformly spaced pilots (i.e. $c_m = mN_s$), the value of ψ is lowered by appropriately choosing the spacing N_s, as demonstrated in Exp. 4. This is related to the well known result that spacing the pilots results in better estimation performance [23, 67]

In Fig. 3.4, we illustrate a more problematic case that can arise in a doubly-selective SISO system, with a single scatterer and path parameter vector $\boldsymbol{\kappa}_1 = (f_1, \tau_1)$ corresponding to the green asterisk in Fig. 3.4.a. We would like the sparsest approximation to correspond to the green triangle, which is the center of the Voronoi cell containing $\boldsymbol{\kappa}_1$. With the addition of noise (constrained to the grey ball), the vector corresponding to the green asterisk may end up closer to the column that corresponds to the red square. The red square is in the vicinity of the desired green triangle column, as shown in Fig. 3.4.b, in spite of the fact that these points are far apart in the parameter space, as shown in Fig. 3.4.a. We say the vector corresponding to the red square is a 'bad neighbor', while 'good neighbors' are those vectors having very similar parameters, as in Fig. 3.4.c.

The designs considered in Sec.'s 2.4.1 and 2.4.2 can result in large values of ψ precisely as shown in Fig. 3.4. In the doubly-selective SISO channel recall the

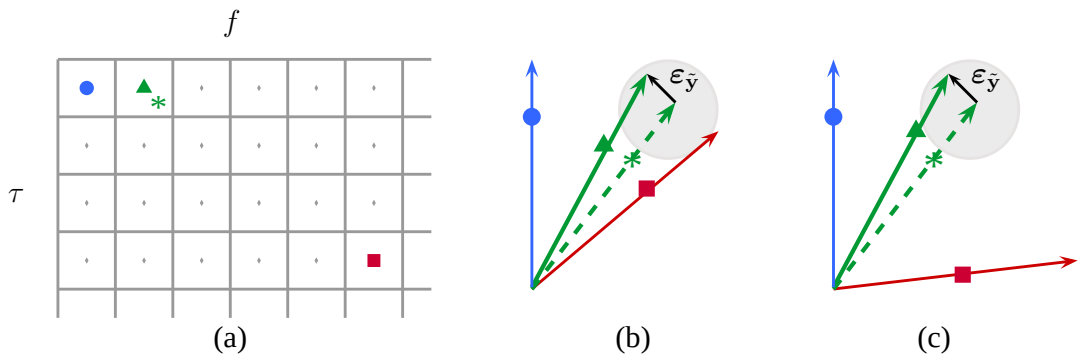

Figure 3.4: Example of quantization of the parameter space (a) where centers of each cell represent a grid point for the parameter space. The green asterisk denotes the true path vector $\boldsymbol{\kappa}_1$. Figures (b) and (c) depict the vectors corresponding to two possible mappings of the parameter space to the vector space spanned by $\boldsymbol{\Upsilon}$.

observations were mapped to a matrix $\boldsymbol{\Upsilon}$, where $\{\mathbf{u}_\mathbf{r}\}_m = e^{j2\pi k(f_{r_f} + \frac{\beta}{KT_s}\tau_{r_\tau})}$. With a finely quantized parameter space, large values of $|\mathbf{u}_\mathbf{r}^H \mathbf{u}_{\mathbf{r}'}|^2$ correspond to grid points indexed by $\mathbf{r} = (r_f, r_\tau)$ and $\mathbf{r}' = (r'_f, r'_\tau)$ such that $|(f_{r_f} - f_{r'_f}) + \frac{\beta}{KT_s}(\tau_{r_\tau} - \tau_{r'_\tau})|^2$ is small, however this does not imply that $\|\mathbf{W}\mathbf{u}_(\boldsymbol{\kappa}_\mathbf{r} - \boldsymbol{\kappa}'_\mathbf{r})\|^2$ is small. For example, if we can find a $\mathbf{u}'_\mathbf{r}$ in the neighborhood of $\mathbf{u}_\mathbf{r}$ such that $(f_{r_f} - f_{r'_f}) \approx -\frac{\beta}{KT_s}(\tau_{r_\tau} - \tau_{r'_\tau})$ where $(f_{r_f} - f_{r'_f})$ is large, the error $\|\mathbf{W}(\boldsymbol{\kappa}_\mathbf{r} - \boldsymbol{\kappa}'_\mathbf{r})\|^2$ can be quite large as well (and similarly $d(\mathbf{r}, \mathbf{r}')$). Hence, not all neighbors result in small error (parameter and/or channel estimation).

This problem is related to the ambiguity of the linear chirp in resolving time and frequency shifts, a problem referred to as range-doppler coupling in radar theory [71]. To overcome this, and attempt to decrease the localized coherence for these cases, it is sufficient to 'move' those problematic vectors $\mathbf{u}_{\mathbf{r}'}$ out of the neighborhood. To accomplish this, we divide the observation vector into two parts $(\mathbf{y}^{(1)}, \mathbf{y}^{(2)})$, each part corresponding to $K/2$ observations and a different pilot design, such that the first part is mapped to a matrix $\{\boldsymbol{\Upsilon}^{(1)}\}_{m, r_\tau + R_\tau r_f} = e^{j2\pi m(f_{r_f} + \frac{\beta}{KT_s}\tau_{r_\tau})}$ while the second part is mapped to

$\{\boldsymbol{\Upsilon}^{(2)}\}_{m,r_\tau+R_\tau r_f} = e^{j2\pi m(f_{r_f}+\frac{\beta}{KT_s}\tau_{r_\tau})}$. Collecting both sets of observations, the overall mapping is

$$\begin{bmatrix} \tilde{\mathbf{y}}^{(1)} \\ \tilde{\mathbf{y}}^{(2)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Upsilon}^{(1)}\boldsymbol{\Delta} \\ \boldsymbol{\Upsilon}^{(2)}\tilde{\boldsymbol{\Delta}} \end{bmatrix} \boldsymbol{\alpha} + \tilde{\mathbf{w}} \tag{3.23}$$

where the matrices $\boldsymbol{\Delta}$ and $\tilde{\boldsymbol{\Delta}}$ are diagonal matrices corresponding to phase parameters (see Appendix A.4) dependent on (f_{r_f}, τ_{r_τ}). If $\mathbf{r}$ and $\mathbf{r}'$ are such that $(f_{r_f} - f_{r'_f}) \approx -\frac{\beta}{KT_s}(\tau_{r_\tau} - \tau_{r'_\tau})$ with large $(f_{r_f} - f_{r'_f})$, the columns $\boldsymbol{\Upsilon}^{(2)}_{\mathbf{r}}$ and $\boldsymbol{\Upsilon}^{(2)}_{\mathbf{r}'}$ will have small correlation, reducing the corresponding contribution to ψ. We remark that although this design allows for a reduction in the localized coherence ψ, the columns of the concatenation of the matrices $\boldsymbol{\Upsilon}^{(1)}\boldsymbol{\Delta}$ and $\boldsymbol{\Upsilon}^{(2)}\tilde{\boldsymbol{\Delta}}$ are still highly correlated. The aim of these designs is not to decrease the mutual coherence, but is instead to control which columns are allowed to be highly correlated, namely those that lead to small error in the channel estimation. The specific signals that lead to such $\boldsymbol{\Upsilon}^{(1)}$ and $\boldsymbol{\Upsilon}^{(2)}$ are given next. Furthermore, similar corrections for the SI inputs for MISO channels follow.

Doubly-Selective Channel

For the doubly-selective SISO channel, we transmit a down-chirp followed by an up-chirp:

$$s(t) = \begin{cases} e^{-j\pi\frac{\beta}{KT_s^2}t^2} & -\tau_{max} \le t < KT_s/2 \\ e^{j\pi\frac{\beta}{KT_s^2}(t-KT_s/2-\tau_{max})^2} & KT_s/2 \le t < KT_s + \tau_{max} \end{cases} \tag{3.24}$$

The vector $\mathbf{y}^{(1)}$ is no different (other than K is now $K/2$) from the one introduced in the proof of Theorem 4.5 in Appendix A.4. Following the same approach, the observations corresponding to the up-chirp portion, after ignoring the samples

corresponding to the transients, can be expressed as:

$$\mathbf{y}^{(2)} = \mathbf{\Delta}_{\mathbf{s}}^{H} \mathbf{\Upsilon}^{(2)} \tilde{\mathbf{\Delta}} \boldsymbol{\alpha} + \mathbf{w} \tag{3.25}$$

where the diagonal matrix $\{\mathbf{\Delta}_{\mathbf{s}}^{H}\}_{k,k} = e^{j\pi\beta\frac{k^2}{K}}$, and the diagonal matrix $\{\tilde{\mathbf{\Delta}}\}_{i(\mathbf{r}),i(\mathbf{r})} = e^{j2\pi f_{r_f}(K/2+\tau_{max})} e^{j\pi\beta\frac{\tau_{r_\tau}^2}{KT_s}}$, after premultiplying $\tilde{\mathbf{\Delta}}^{-1}\mathbf{y}^{(2)}$ we have the desired form with:

$$\begin{aligned} \{\mathbf{\Upsilon}^{(1)}\}_{k,r_\tau+R_\tau r_f} &= e^{j2\pi k(f_{r_f}+\frac{\beta}{KT_s}\tau_{r_\tau})} \\ \{\mathbf{\Upsilon}^{(2)}\}_{k,r_\tau+R_\tau r_f} &= e^{j2\pi k(f_{r_f}-\frac{\beta}{KT_s}\tau_{r_\tau})} \end{aligned}$$

MISO Time-Selective Channel

The designs proposed in Theorem 2.4.6 resulted in a diagonal training matrix $\mathbf{S} = \mathbf{\Delta}_{\tilde{\mathbf{s}}}$ (for simplicity assume $\mathbf{S} = \mathbf{I}$). The first $k = 0, .., N_{tx} - 1$ observations were mapped to a matrix $\{\mathbf{\Upsilon}^{(1)}\}_{k,i(\mathbf{r})} = e^{j2\pi k(f_{r_f}+\mu_{r_\mu})}$. We propose the reuse of the antennas in the reverse order in the next K observations, such that $\mathbf{S} = \mathbf{J}$ where $\mathbf{J}$ is a matrix with ones on the anti-diagonal and zeros elsewhere. This next set of observations:

$$\mathbf{y}^{(2)} = \sum_{\mathbf{r}} e^{j2\pi f_{r_f} K} \mathbf{\Omega}_{r_f} \mathbf{J} \mathbf{e}_{r_\mu} \{\boldsymbol{\alpha}\}_{i(\mathbf{r})} + \mathbf{w} \tag{3.26}$$

where $e^{j2\pi f_{r_f} K}$ is the phase shift in the Doppler as we are examining the next set of $K = N_{tx}$ observations. Noting that $\mathbf{J}\mathbf{e}_{r_\mu} = \mathbf{e}_{-r_\mu} e^{j2\pi\mu_{r_\mu}(N_{tx}-1)}$, using the same approach as in Appendix A.5, the observations $\mathbf{y}^{(2)} = \mathbf{\Upsilon}^{(2)}\tilde{\mathbf{\Delta}}\boldsymbol{\alpha}$ where the diagonal matrix $\{\tilde{\mathbf{\Delta}}\}_{i(\mathbf{r}),i(\mathbf{r})} = e^{j2\pi(f_{r_f}K+\mu_{r_\mu}(N_{tx}-1))}$. We now have the desired form:

$$\begin{aligned} \{\mathbf{\Upsilon}^{(1)}\}_{k,i(\mathbf{r})} &= e^{j2\pi k(f_{r_f}+\mu_{r_\mu})} \\ \{\mathbf{\Upsilon}^{(2)}\}_{k,i(\mathbf{r})} &= e^{j2\pi k(f_{r_f}-\mu_{r_\mu})} \end{aligned}$$

MISO Frequency-Selective Channel

For the designs in Theorem 2.4.7, the v-th antenna, $v = 0, ..., N_{tx}-1$, transmitted a frequency $s_v(t) = e^{-j2\pi\frac{vt}{KT_s}}$. After projecting over these frequencies at the receiver the observations were mapped to a matrix $\{\boldsymbol{\Upsilon}^{(1)}\}_{v,i(\mathbf{r})} = e^{j2\pi v(\frac{\tau_{r_\tau}}{KT_s}+\mu_{r_\mu})}$. We propose the reuse of the antennas by transmitting the conjugate frequencies in the second block of training, i.e. $s_v(t) = e^{j2\pi\frac{vt}{KT_s}}$. Following the same steps in Appendix A.6, the observations:

$$\mathbf{y}^{(2)} = \mathbf{P}\sum_{\mathbf{r}} \boldsymbol{\Omega}_{r_\tau}\mathbf{e}_{r_\mu}\{\boldsymbol{\alpha}\}_{i(\mathbf{r})} + \mathbf{w} \tag{3.27}$$

where $\{\mathbf{P}\}_{k,v} = e^{j2\pi kv/K}$ and the diagonal matrix $\{\boldsymbol{\Omega}_{r_\tau}\}_{v,v} = e^{-j2\pi v\tau_{r_\tau}/(KT_s)}$. Thus after premultiplying $\mathbf{y}^{(2)}$ by $K^{-1}\mathbf{P}^H$ the observations are mapped to a matrix $\{\boldsymbol{\Upsilon}^{(2)}\}_{v,i(\mathbf{r})} = e^{j2\pi v(\frac{-\tau_{r_\tau}}{KT_s}+\mu_{r_\mu})}$. We now have the desired form:

$$\begin{aligned}\{\boldsymbol{\Upsilon}^{(1)}\}_{v,i(\mathbf{r})} &= e^{j2\pi v(\frac{\tau_{r_\tau}}{KT_s}+\mu_{r_\mu})} \\ \{\boldsymbol{\Upsilon}^{(2)}\}_{v,i(\mathbf{r})} &= e^{j2\pi v(\frac{-\tau_{r_\tau}}{KT_s}+\mu_{r_\mu})}\end{aligned}$$

3.6 Experiments

In this section we present several experiments using the modified designs presented in Sec. 3.5. Recall, in Sec. 2.3.1, a tractable relaxation to (2.13), known as basis pursuit, was presented. In the presence of noise and modeling error, the basis pursuit method (2.14) can be modified by placing a constraint on the reconstruction error such as [19]:

$$\hat{\boldsymbol{\alpha}} = \arg\min \|\boldsymbol{\alpha}\|_1 \;\text{ s.t. }\; \|\tilde{\mathbf{y}} - \boldsymbol{\Upsilon}\boldsymbol{\alpha}\|_2 \leq \eta \tag{3.28}$$

Similar approaches for estimating sparse $\boldsymbol{\alpha}$ in the presence of noise, include [22] which replaces the inequality constraint with $\|\boldsymbol{\Upsilon}^H(\mathbf{y} - \boldsymbol{\Upsilon}\boldsymbol{\alpha})\|_\infty \leq \eta$ and is known

as the the *Dantzig Selector*, or allowing some remaining residual error in matching pursuit methods.

In the following, we first present some similar estimation schemes that will be used for comparison in the experiments. This is followed by several experiments where we use (3.28) to estimate the channel. It is implemented using a log-barrier interior point method where we set $\eta = (C_\eta \log(R)\sigma_w^2)^{1/2}$ where $C_\eta \approx 2$.

3.6.1 Estimation and MSE Performance of Competing Schemes

We discuss similar methods of channel estimation and their corresponding mean-squared error (MSE) performance, $\mathbb{E}\|\hat{\mathbf{h}} - \mathbf{h}^\star\|^2$, under the assumption that no modeling error is present i.e. $\tilde{\mathbf{y}} = \mathbf{P}\boldsymbol{\mathcal{X}}\mathbf{h}^\star + \mathbf{w}$. As discussed in the introduction, BEM is often used for channels having memory and time-variations. For a set of K observations, these methods model the time-variations of the channel over a subspace of $\mathbb{C}^K$ using time-varying basis functions (e.g. [16], [39]), $\mathbf{h}_{\text{BEM}} = (\boldsymbol{\Gamma} \otimes \mathbf{I}_L)\mathbf{a}^\star$ where $\boldsymbol{\Gamma}$ consists of the time-varying basis functions (e.g. polynomial basis functions [16], exponential basis functions [39], or Karhunen-Loeve basis functions [96]). This is similar to the model (2.6), where in BEM, the matrix $\boldsymbol{G} = \boldsymbol{\Gamma} \otimes \mathbf{I}_L$ typically has full column rank. The least squares estimate using a basis expansion model and its associated MSE are given by:

$$\hat{\mathbf{h}}_{\text{BEM}-\text{LS}} = \boldsymbol{G}(\underbrace{\mathbf{P}\boldsymbol{\mathcal{X}}\boldsymbol{G}}_{\boldsymbol{\Upsilon}})^\dagger \tilde{\mathbf{y}} \tag{3.29}$$

$$\text{MSE}_{\text{BEM}-\text{LS}} = \sigma_w^2 \text{trace}\left((\boldsymbol{\Upsilon}^\dagger)^H \boldsymbol{G}^H \boldsymbol{G} \boldsymbol{\Upsilon}^\dagger\right) + \|\boldsymbol{G}(\boldsymbol{\Upsilon}^\dagger\boldsymbol{\Upsilon} - \mathbf{I})\mathbf{a}^\star\|^2 \tag{3.30}$$

Another approach to channel estimation uses a parametric model for the chan-

nel, as in (2.5), and attempts to estimate the path parameters $\boldsymbol{\kappa} = (\boldsymbol{\kappa}_1, ..., \boldsymbol{\kappa}_Q)$, $\mathbf{a}$, and Q in order reconstruct the channel estimate from these parameters: $\hat{\mathbf{h}}_{\mathrm{PM}} = \mathbf{G}(\hat{\boldsymbol{\kappa}})\hat{\mathbf{a}}$ where the q-th column of $\mathbf{G}(\boldsymbol{\kappa})$ can be defined via (2.5) as $[\mathbf{G}(\boldsymbol{\kappa})]_q = \mathbf{e}(\nu_q) \otimes \mathbf{e}(\mu_q) \otimes \mathbf{e}(f_q) \otimes \mathbf{p}(\tau_q)$. Under the assumption that Q is known a priori, the estimation of the parameters $(\boldsymbol{\kappa}, \mathbf{a})$ may be determined via maximum likelihood (ML) estimation resulting in:

$$(\hat{\boldsymbol{\kappa}}, \hat{\mathbf{a}}) = \arg \min_{(\boldsymbol{\kappa}, \mathbf{a})} \|\tilde{\mathbf{y}} - \boldsymbol{\Upsilon}(\boldsymbol{\kappa})\mathbf{a}\|_2^2 \tag{3.31}$$

where $[\boldsymbol{\Upsilon}(\boldsymbol{\kappa})]_q = [\mathbf{P}\boldsymbol{\mathcal{X}}\boldsymbol{G}(\boldsymbol{\kappa})]_q$. We refer to this approach as the *parametric maximum likelihood estimate* (PM-ML). The estimation of these parameters is complicated due to the nonlinear dependence on the parameters $\boldsymbol{\kappa}$, but the advantage of the ML estimate is that is asymptotically efficient. The MSE, if Q is known a priori, may be lower bounded via the Cramer Rao Lower Bound (CRLB). The approach of quantizing the parameter space and using sparse estimation algorithms, in comparison with the PM-ML approach, does not require exact knowledge Q, but instead requires an upperbound on Q in order to design the pilot signals. For this reason we cannot use $\mathrm{MSE}_{\mathrm{PM-ML}}$ as a lower bound for the MSE of sparse estimation algorithms, as the parameter space of the PM-ML is neither discrete nor sparse, and the sparse estimation approach is biased. This being said, it still provides a useful comparison for the sparse estimation methods.

3.6.2 Numerical Experiments

Experiment 1. As a motivating example, we implement the proposed scheme on a single realization of the doubly selective SISO channel considered in Sec. 2.4.1, with spreading factor $\tau_{\max}\xi_{max} \approx 1$. We define $L = 12$ as the approximate number of taps in the discrete-time channel response, and $f_{max} = 1/12$, observing the

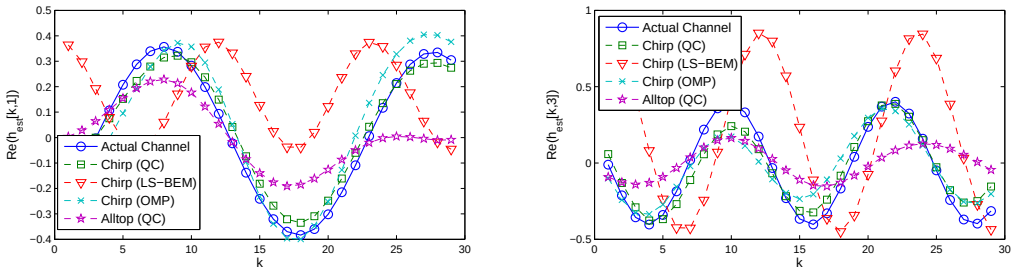

Figure 3.5: Channel estimates on two taps for a realization of SISO doubly-selective channel with $f_{max} = 1/12$ and $\tau_{max}/T_s = 12$, and $Q = 5$ paths with $SNR = 25$ dB.

system over $K = 29$ samples. The system consists of $Q = 5$ scatterers with continuous valued path parameters generated uniformly at random, $f_q \sim \mathcal{U}(0, 1/12)$, $\tau_q/T_s = \mathcal{U}(0, L-1)$ and $a_q \sim \mathcal{CN}(0, 1/Q)$. The transmit signal is generated using a digitally-modulated signal $s(t) = \sum_{n=0}^{N-1} p(t - nT_s)x[n]$ where $p(t) = \text{sinc}(\pi t/T_s)$ with finite duration of 8 symbol periods and a cyclic-prefix is assumed. The system is generated using (2.2).

We compare the performance of our down-up chirp design with the design in [45], and a LS-BEM method (3.29). The design proposed in [45] uses an Alltop sequence $x[n] = e^{j2\pi n^3/N}$ and a uniform grid for the parameters $\epsilon_\tau = T_s$ with $R_\tau = K$ for the delays and $\epsilon_f = 1/K$ and $R_f = K$ for the normalized Doppler shifts. The down-up chirp proposed in Sec. 3.5 is approximated with $x[n] = e^{-j\pi\frac{\beta n^2}{N}}$ where we have set $\beta = 2(1 - f_{max})$ for the first 14 symbols while using the up-chirp on the remaining 15 samples. The uniformly spaced grid points for the down-up chirp design are $\epsilon_\tau = T_s(L-1)/(2L-1)$ with $R_\tau = 2L$ and $\epsilon_f = 1/(12K)$ with $R_f = 40$. We use both (3.28) (denoted as QC in Fig. 1) and also Orthogonal Matching Pursuit (denoted as OMP in Fig. 1) for estimating the channels in these designs. For the LS-BEM estimate, as $K = 29$ and $L = 12$, we are limited

to two time-varying basis functions, and we use complex exponential vectors with normalized frequencies 0 and 1/12 while still using the down-up chirp signal.

In Fig. 3.5 we plot the real part of two taps for this channel realization. As expected, the down-up chirp and its corresponding grid appear to outperform the other methods for this realization. Further, the LS-BEM, with only two basis functions modeling the time-variations of the channel, is not able to accurately model the time variations. For this particular realization, the corresponding normalized squared error SE $= (1/K)\|\hat{\mathbf{h}} - \mathbf{h}\|_2^2$ is $\text{SE}(\text{Chirp}_{\text{QC}}, \text{Chirp}_{\text{OMP}}, \text{Chirp}_{\text{LS-BEM}}, \text{Alltop}_{\text{QC}}) = (0.095, 0.201, 4.93, 0.462)$.

Experiment 2. We again consider the doubly-selective SISO channel considered in Sec. 2.4.1 with $K = 17$ observations and assuming $L = 5$. The Q channel parameter sets are drawn from a uniform distribution over a continuous interval, i.e. $f_q \sim \mathcal{U}(-.005, .005)$ and $\tau_q \sim \mathcal{U}(0, L-1)$, while the path attenuation for each of the Q paths is generated i.i.d. as $a_q \sim \mathcal{CN}(0, 1/Q)$. The received observations are generated using the continuous parameter model in (2.2) where again $s(t) = \sum_{n=0}^{N-1} x[n]\text{sinc}(\pi(t - nT_s))$ with $N = 17$ symbols. We compare the channel MSE for our proposed design and grid, averaging over 1000 channel realizations, with the Alltop sequence $x[n] = e^{j2\pi n^3/N}$ considered in [45] and the random binary single-carrier design in [9].

The grid parameters for the Alltop sequence (as suggested in [45]) are $\epsilon_\tau = T_s$ and $R_\tau = L$ for the delays and $\epsilon_f = 1/K$ and $R_f = K$ for the normalized Doppler shifts. The same grid parameters are used for random binary training sequence proposed in [9]. The down-up chirp design sets $x[n] = e^{-j\pi\frac{\beta n^2}{N}}$ with $\beta = 2(1-f_{max})$ for the first $N = 8$ symbols and uses the up-chirp for the remaining 8 symbols. The grid parameters with this down-up chirp are $\epsilon_\tau = T_s(L-1)/(2L-1)$ with

$R_\tau = 2L$, and $\epsilon_f = f_{max}/K$ with $R_f = K$. We remark that this choice of grid parameters for the down-up chirp results in a finer grid compared to that proposed in [45]. In Fig. 3.6, we compare the MSE of the channel estimate normalized by the number of observations K using (3.28). We also plot the corresponding CRLB for the corresponding PM-ML estimate (3.31) as a comparison, assuming Q is known apriori.

We observe that the chirp sequence outperforms the Alltop sequence as well as the random binary sequence. For the parameter grids specified above, the Alltop and binary sequences have normalized frequency spacing of $1/K$ and delay spacing of T_s seconds, while the parameter grid specified for the chirp sequence has frequency spacing $1/(100K)$ and delay spacing $T_s(L-1)/(2L-1)$ seconds. We observe the grid with fine spacing outperforms the one with coarse spacing. In Fig. 3.6, we also plot the MSE of the Alltop sequence using the same grid as the down-up chirp. Interestingly, when the same grid is used for both sequences, both the chirp and Alltop sequences have very similar performance, confirming that a higher resolution grid does not hurt performance (though it does increase the complexity of the recovery algorithm). We note that sensing matrix $\boldsymbol{\Upsilon}$ has mutual coherence close to unity for the fine grid. The similar performance of the down-up chirp sequence and the Alltop sequence can be explained by noting that their instantaneous frequencies $\dot{\phi}(t)$ have a similar trend in the time-frequency plane, since for the first $\dot{\phi}(t) \approx a'|t-a|$ and for the second $\dot{\phi}(t) \approx b'(t-b)^2$.

As observed in Fig. 3.6 and later experiments, the MSE appears to saturate at high SNR. There are two reasons for this behavior; the first is a result of modeling error, while the second is the choice of η in (3.28). Obviously, if η is too large or small our performance may greatly suffer.

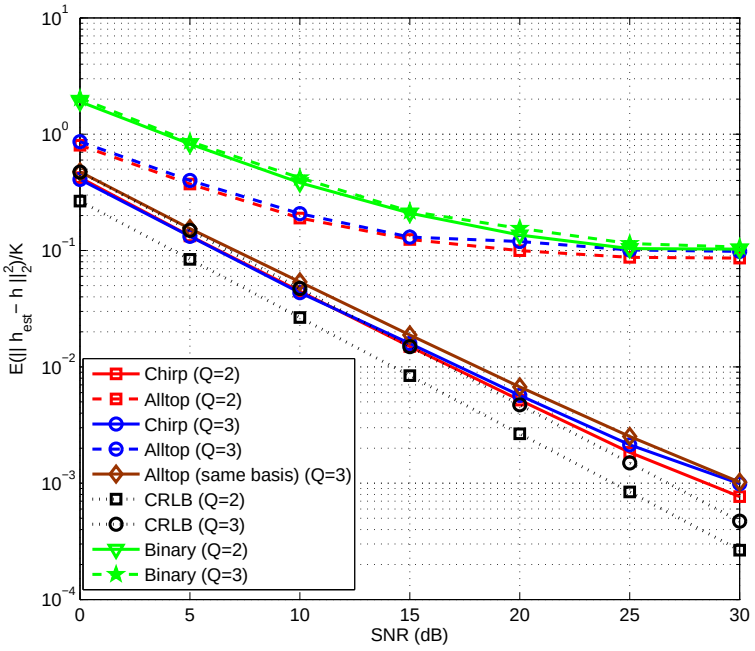

Figure 3.6: Channel estimate MSE of SISO doubly-selective channel for both chirp sequence and Alltop sequence [45]

Experiment 3. In this experiment, we consider a SISO frequency-selective channel and approximate the SI design in Sec. 2.4.1 using OFDM. In Fig. 3.7, we plot the MSE of the channel frequency response for a system consisting of $K = 2048$ carriers, a cyclic prefix (CP) of length $N_{cp} = L = 512$, as done in [87], averaging over 30 iterations. We assume $Q = 3$ scatterers with continuous delays generated uniformly at random, $\tau_q/Ts \in (0, N_{cp} - 1)$, and path gains $a_q \sim \mathcal{CN}(0, 1/Q)$. We compare the method used in [87] (denoted as [TH] in Fig. 3.7), with the approach of modeling the channel with an overcomplete basis $\mathbf{h} = \boldsymbol{G\alpha}$ as in Prop. 2.2.1 and using (3.28) to estimate $\boldsymbol{\alpha}$ (denoted as Sparse Est.). The least-squares estimate (LS) is also included for comparison.

In the LS estimate the $M = 512$ pilots are uniformly spaced every $K/N_{cp} = 4$ carriers, while in Sparse Est. and [TH], the $M = 256$ pilots are uniformly selected

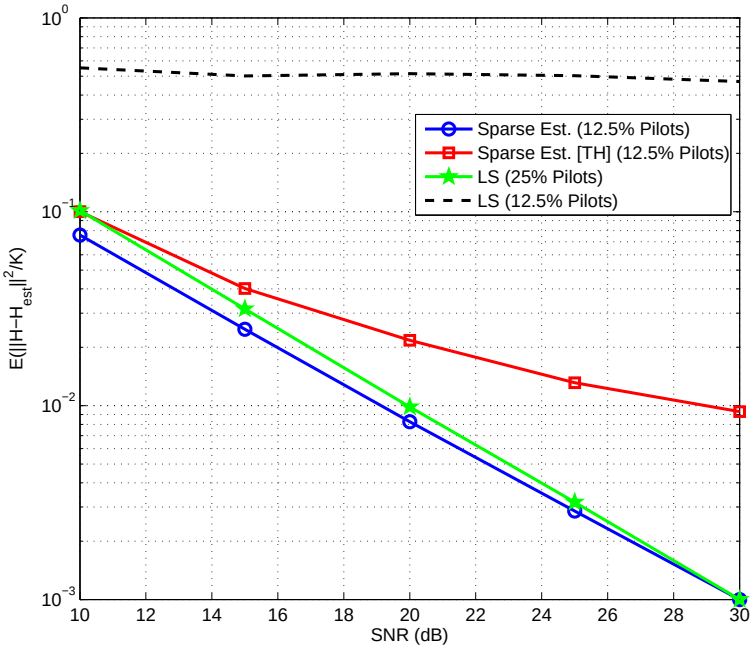

Figure 3.7: Channel MSE of SISO frequency-selective channel for $K = 2048$, $M = 256$, $N_{cp} = 512$ and $Q = 3$ scatterers.

at random among those used for the LS estimate. When modeling the channel as $\mathbf{h} = \boldsymbol{G}\boldsymbol{\alpha}$, the $L \times R_\tau$ matrix $\{\boldsymbol{G}\}_{l,r_\tau} = \text{sinc}(\pi(l - r_\tau L/R_\tau)$. The method considered in [87] corresponds to $R_\tau = L$ (such that $\mathbf{h} = \boldsymbol{\alpha}$), while for Sparse Est. we use $R_\tau = 2L$. In Fig. 3.7 we plot the MSE of the channel frequency response estimate $\hat{\mathbf{H}} \triangleq \mathbf{P}\hat{\mathbf{h}}$ where the $K \times L$ DFT matrix $\{\mathbf{P}\}_{k,l} = e^{-j2\pi kl/K}$. As observed, a finer quantization of the parameter space allows for a reduction in the MSE of the channel.

In Fig. 3.8 we compare the channel MSE of the OFDM system for various values of R_τ, using $K = 128$ carriers, a CP length $N_{cp} = L = 32$, and $M = 16$ pilots generated in the same manner as in Fig. 3.7. We assume $Q = 2$ scatterers with continuous delays generated uniformly at random $\tau_q/T_s \in (0, N_{cp} - 1)$. We observe the MSE performance improves as R_τ increases, corresponding to a decrease in the

quantization error in the parameter space. We remark that the method considered in [87] corresponds to $R_\tau = L$, and observe that it is not capable of accurately estimating the channel in this setting. We remark that the method proposed [88] is equivalent to that in [87] in this setting, as we assume the Doppler shifting is negligible. Similar to the previous experiments, we remark that the a finer grid does increase the complexity of the recovery algorithm.

In Fig. 3.9 we examine the MSE for various M and Q fixing $R_\tau = 4L$. We observe that the number of pilots necessary for accurate estimation of the channel in the presence of noise and modeling error is greater than the number derived for sufficiently informative inputs, however, we also observe that a finer quantization allows for a reduction in the number of pilots. This number of pilots used is still much less than the number of pilots needed to obtain small restricted isometry constants and mutual coherence.

Experiment 4. In this experiment, we consider the SISO time-selective channel in Sec. 2.4.1, demonstrating the well known result that spacing the pilots results in better localized coherence ψ and channel MSE. We consider a maximum normalized Doppler shift $f_{max} = 1/8$ and a uniform grid for the parameter space consisting of $R_f = 60$ points with spacing $\epsilon_f = (8(R_f - 1))^{-1}$. We assume a total of $M = 10$ uniformly spaced pilots, i.e. $c_m = mN_s$.

To compute ψ, we set $a = \sigma_a^2 = 1/2$ (as we assume $Q = 2$ paths) and define $d(\mathbf{r}, \mathbf{r}') = \|(\mathbf{I} - \boldsymbol{G}_{\mathbf{r}'}\boldsymbol{\Upsilon}_{\mathbf{r}'}^{\dagger}\mathbf{P}\boldsymbol{\mathcal{X}})\boldsymbol{G}_{\mathbf{r}}\|2$ where $\mathbf{r} = r_f$ and the length K column vector $\{\boldsymbol{G}_{\mathbf{r}}\}_k = e^{j2\pi k r_f \epsilon_f}$. The matrix $\mathbf{P}\boldsymbol{\mathcal{X}}$ is simply an $M \times K$ selection matrix corresponding to the pilot locations, namely $(\mathbf{P}\boldsymbol{\mathcal{X}})_{m,k} = \delta[k - mN_s]$. Each vector $\boldsymbol{G}_{\mathbf{r}}$ corresponds to a different basis function for the time-variations of the channel over $K = 80$ observations of which $M = 10$ are pilots. We denote $\psi^{(1)}$ as the

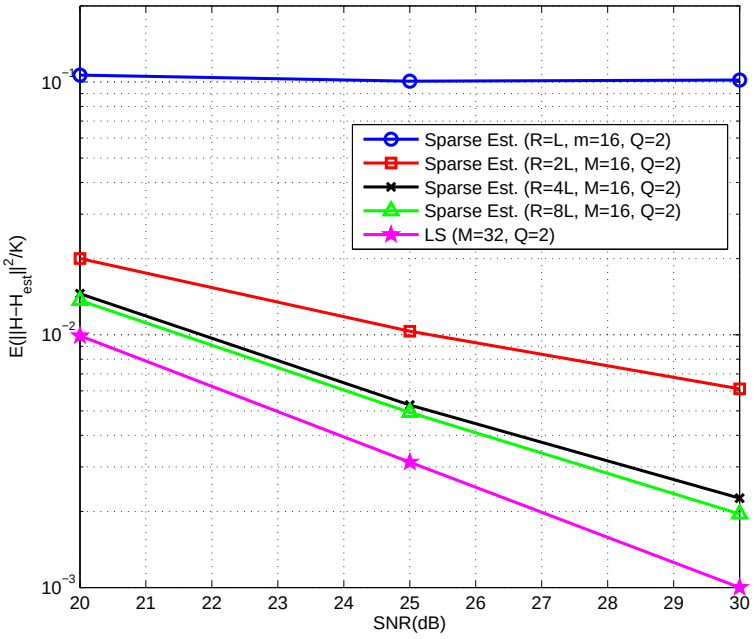

Figure 3.8: Channel MSE for OFDM system with $K = 128$ carriers, $M = 16$ pilots, CP length $N_{cp} = 32$ and $Q = 2$ scatterers.

localized coherence using $d(\mathbf{r}, \mathbf{r}')$ defined above, while $\psi^{(2)}$ is the localized coherence computed using $d(\mathbf{r}, \mathbf{r}') = \|(r_f - r'_f)\epsilon_f\|_2$. The values of localized coherence ψ and mutual coherence μ_c are shown in Table 3.1 for different pilot spacings. As expected, the localized coherence decreases for increasing N_s for both definitions of the error $d(\mathbf{r}, \mathbf{r}')$, while the mutual coherence for all spacings is approximately one. While the largest possible spacing for the pilots is $N_s = K/M = 8$, we cannot choose this spacing exactly because it will potentially undersample the Doppler frequencies. Indeed the mutual coherence for $N_s = 8$ is exactly 1. There is no system cost in increasing N_s other than the latency of the receiver and the only cost in decreasing ϵ_f is in the complexity of the recovery algorithm.

In Fig. 3.10 we plot the normalized MSE of the time-varying channel $\mathbb{E}\{\|\mathbf{h}^\star - \hat{\mathbf{h}}\|^2\}/K$ consisting of $Q = 2$ Doppler shifts for various values of N_s using (3.28)

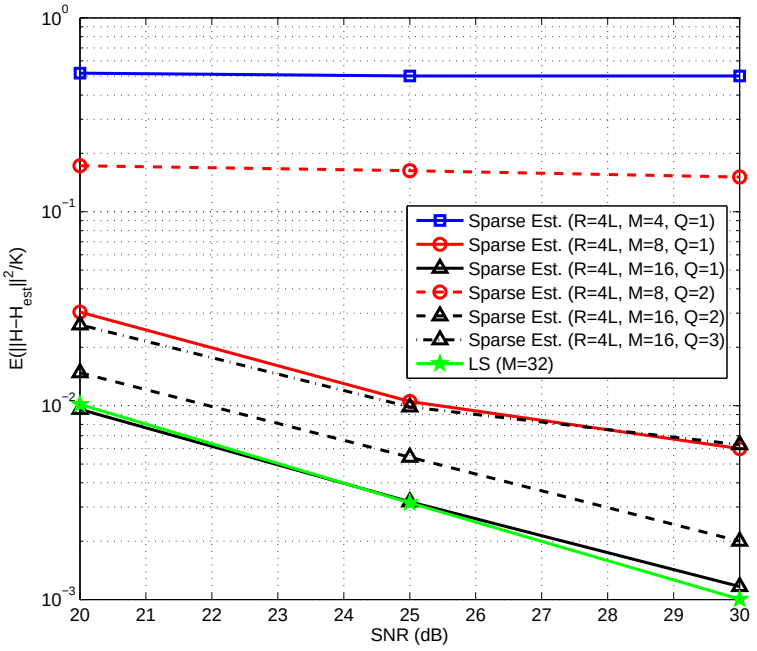

Figure 3.9: Channel MSE for OFDM system with $K = 128$ carriers, CP length $N_{cp} = 32$, for various Q and M .

Table 3.1: Localized Coherence and Mutual Coherence for Exp. 4.

N_s	1	2	3	4	5	6	7
$\psi^{(1)}$	39.6	22.3	16.0	12.9	11.3	10.4	9.9
$\psi^{(2)}$	0.708	.512	.467	.450	.441	.436	.434
μ_c	0.99	0.99	0.99	0.99	0.98	0.97	0.96

to estimate $\boldsymbol{\alpha}$. The MSE is averaged over 500 iterations and in each iteration the normalized Doppler shifts f_q are generated uniformly at random in $(-1/16, 1/16)$ and $a_q \overset{i.i.d.}{\sim} \mathcal{CN}(0, 1/Q)$. As predicted by the localized coherence, spacing of the pilots allows a much better estimate of the time-varying channel.

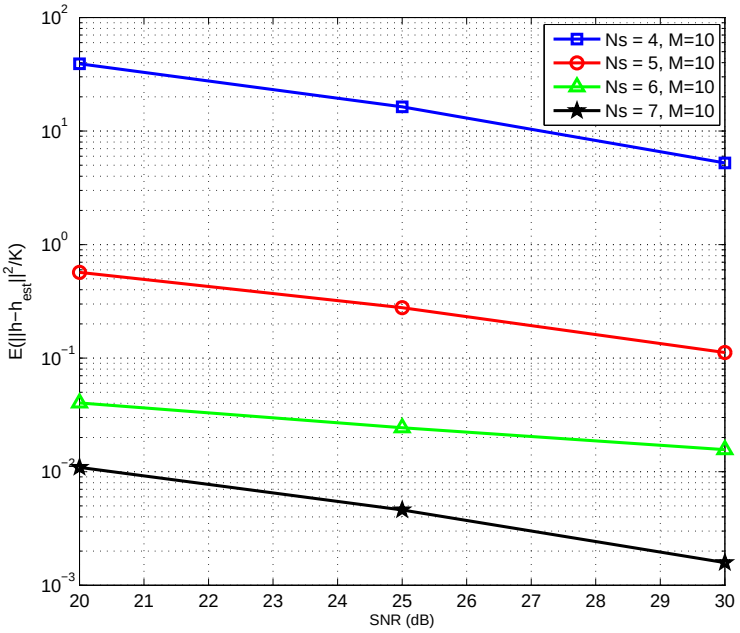

Figure 3.10: Channel MSE for SISO time-selective channel with $Q = 2$ paths with various spacing of $M = 10$ pilots $c_m = mN_s$.

3.7 Summary

In this chapter, we analyzed the performance of an ideal sparse recovery algorithm when applied to the problem of channel estimation. To estimate the channel impulse response, we constructed an overcomplete basis for the channel impulse response by quantizing the parameter space of delays, Doppler shifts, and angles. For an appropriate choice of the allowable reconstruction error in the observations, we observed the most probable sets of basis functions returned by the ideal sparse recovery algorithm corresponded to those sets of minimal size that closely spanned the channel space. Based on this observation, for a fixed quantization of the channel parameter space, we introduced a metric called *localized coherence*, that can be used to compare input signal designs in terms of their channel estimation performance. Using the metric of localized coherence to the SI input designs discussed in Sec.

2.4, we proposed modifications to obtain better channel estimation performance.

We observe through simulation that in practical situations where the parameters of the channel (delays, Doppler shift, and angles of departure) do not lie on the grid corresponding to the quantized parameter space, a fine grid is preferred over a coarse grid. Further, we observed, through simulation, that the metric of localized coherence provides a way to compare training signal designs in terms of channel estimation performance, even when using suboptimal sparse recovery methods such as OMP and basis pursuit in place of the ideal sparse recovery algorithm analyzed.

CHAPTER 4

DIVERSITY FOR RANDOMIZED COOPERATION IN ASYNCHRONOUS DISPERSIVE LINKS

4.1 Motivation and Related Work

As we indicated in the introduction, our work on channel estimation is motivated by the need for advanced methods to estimate the distortion resulting from cooperative schemes. Ideally, cooperative schemes bring advantages that are analogous to those of MIMO systems [53, 77]. These systems combat the effects of fading with spatial diversity.

In [82], the authors introduced the idea of randomized cooperation for fully decentralized cooperative protocols and analyzed the diversity attainable over flat-fading channels. In this cooperative scheme, each node is unaware of the effective code being employed by other users, as well as the number of users T cooperating. The randomized coding rule targets a fixed diversity order of D, which is independent of the number of nodes actually cooperating.

The motivation behind this work is that nodes cooperating without central control will encounter the practical problem of synchronizing their access to the channel. Their imperfect source-clock acquisition and relative displacement will cause difference in delays easily on the order of microseconds, giving rise to linear distortion. Assuming that the channel can be estimated, delay and carrier offset impact the system performance in a different manner.

In fact, a random transmission delay can be the source of diversity at the receiver [98]. The goal of this chapter is to investigate what class of randomized

coding techniques can more easily harvest delay diversity. Hence, for the sake of the theoretical analysis the effect of carrier offset is considered sufficiently mild so that it leads to a simple block-fading model.

Through simulation, we compare and analyze several different coding schemes for this randomized coding method. These include Alamouti space-time coding (STC) [5], Time-Reversal STBC (TR-STBC) [59], and space-frequency OFDM (SF-OFDM) [56]. In addition to comparing these different coding schemes, we also simulate the performance of these schemes in the presence of an doubly-selective HF channel, using the well known Watterson model [97].

Other cooperative transmission schemes that are applicable to a decentralized scenario include [98] and [74]. In [98] asynchronous cooperative diversity was considered, where random delays are introduced by cooperating nodes, and our scheme includes this architecture as a special case. The scheme in [99] has a similar formulation to our scheme, where each node transmits the product of a space-time code matrix with a preassigned vector code but, as a result, a preliminary code allocation phase is necessary. In [58], the authors considered the effects of asynchrony among cooperative nodes in a distributed scenario. However, the channels were assumed time-invariant over each packet transmission, but randomly varying between packets. Further The proposed transmission scheme still required coordination between nodes in order to handle the timing offsets resulting from asynchrony.

4.2 System Model

We consider a network with T cooperating nodes sharing a common message $\mathbf{s} = (s_0, ..., s_{M-1})^T$, which is the perfect estimate of a message previously transmitted by a source and the nodes need to relay (the strategy is, thus, a *decode and forward* method [54]). The cooperating nodes are recruited on the fly based on, for example, the power received from the original source. We assume the nodes are unaware of other cooperating nodes, have acquired the source clock and time of arrival of the source preamble sequence, and make their best attempt to retransmit in unison in the following time slot given their source timing estimate.

The cooperating nodes generate their codes by forming a predetermined $K \times D$ code matrix $\mathbf{G}(\mathbf{s})$ (D is the maximum diversity order of the underlying space-time code and K is its duration). In a centralized scheme, a generic column $\mathbf{x}_d \in \mathbf{G}(\mathbf{s})$ is assigned to a specific node. In a randomized scheme [82], instead, the node's code is generated locally as $\mathbf{G}(\mathbf{s})\mathbf{r}_i$ where $\mathbf{r}_i$ is an $D \times 1$ random vector. To capture the lack of synchronization among the cooperating nodes, we assume the channel between the i-th node and the destination is a multi-path channel with maximum finite memory L and impulse response $\mathbf{h}_i$. The transmission is structured in blocks which are interleaved with guard intervals, with at least L zeros (the use of a cyclic prefix is also possible), to eliminate inter-block interference. The received signal is the summation of the T transmissions from the cooperating nodes

$$\mathbf{y} = \sum_{i=1}^{T} \mathbf{H}_i \mathbf{G}(\mathbf{s}) \mathbf{r}_i + \mathbf{w} = \sum_{d=1}^{D} \left(\sum_{i=1}^{T} \mathbf{H}_i r_i(d) \right) \mathbf{x}_d + \mathbf{w} \tag{4.1}$$

where $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, N_o\mathbf{I})$ and the $(K + L) \times K$ matrix[1] $\mathbf{H}_i = \mathcal{T}(\mathbf{h}_i)$ where $\mathcal{T}(\mathbf{h}_i)$ denotes the Toeplitz convolution matrix formed from the vector $\mathbf{h}_i$. Letting $\mathbf{h} =$

[1]Here, we assume a linear time-invariant channel with impulse response $\mathbf{h}_i$.

$(\mathbf{h}_1^T, ..., \mathbf{h}_T^T)^T$, we rearrange the terms in (4.1) to obtain:

$$\begin{aligned} \mathbf{y} &= \sum_{d=1}^{D} \mathcal{T}(\mathbf{x}_d) \left[(r_1(d), ..., r_T(d)) \otimes \mathbf{I}_{L+1} \right] \mathbf{h} + \mathbf{w} \\ &= \boldsymbol{\mathcal{X}} (\mathbf{R} \otimes \mathbf{I}_{L+1}) \mathbf{h} + \mathbf{w} \end{aligned} \tag{4.2}$$

where $\boldsymbol{\mathcal{X}} = (\mathcal{T}(\mathbf{x}_1), ..., \mathcal{T}(\mathbf{x}_D))$ and $\mathbf{R} = (\mathbf{r}_1, ..., \mathbf{r}_T)$. From (4.2) we observe that the delay dispersion can be mapped into a specific block Toeplitz structure, and the randomization leads to an *effective channel* $\tilde{\mathbf{h}} = (\tilde{\mathbf{h}}_1^T, ..., \tilde{\mathbf{h}}_D^T)^T = (\mathbf{R} \otimes \mathbf{I}_{L+1})\mathbf{h}$. Hence, decoders already designed for multiple-antenna space-time codes can be directly used for randomized space-time coding. Further, the complexity of the channel state estimation is reduced when $D < T$ since the effective channel $\hat{\mathbf{h}}$ is shorter than the actual channel $\mathbf{h}$.

With an effective channel vector of dimension $D(L+1)$ the average error probability may scale at most $O(SNR^{-D(L+1)})$. As argued in [85], a necessary condition for this trend is that the space time code designs meet the following criterion:

C1) *Rank criterion for* $\boldsymbol{\mathcal{X}}$: For any pair of code matrices $\{\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{X}}_j\}$, the matrix $(\boldsymbol{\mathcal{X}}_i - \boldsymbol{\mathcal{X}}_j)$ has full column rank $\forall i \neq j$, i.e., of rank $D(L+1)$.

Lemma 4.2.1. *Under the assumptions leading to the model in (4.2) the rank condition C1) is met i.f.f.*

C2) *The code sequence* $\mathbf{x}_d$ *has length* K *such that* $K \geq (D-1)L + D$

C3) *For any two distinct source messages* $\mathbf{s}_i, \mathbf{s}_j \in \mathcal{M}$ *the polynomials formed by taking the Z-transforms of the columns of the matrix* $\mathbf{A}_{ij} = \boldsymbol{\mathcal{X}}_i - \boldsymbol{\mathcal{X}}_j = (\mathbf{a}_{ij}^{(1)}, \ldots, \mathbf{a}_{ij}^{(D)})$ *are co-prime.*

Condition C2) simply states that the matrix $\mathbf{A}_{ij}$ is tall or at least square in order to have full column rank. Condition C3) implies that all the columns of $\mathbf{A}_{ij}$ are

linearly independent and is equivalent to conditions invoked in several second-order blind channel identification methods [66]. The question of the spectral efficiency one can obtain while satisfying the rank criterion is beyond the scope of this thesis, but nothing appears to prevent the existence of such codes.

In the following, let $\mathcal{M} = \{\mathbf{s}_1, ..., \mathbf{s}_{|\mathcal{M}|}\}$ denote the message set with each message equally likely. We assume the effective channel $\hat{\mathbf{h}}$ is known at the destination. Let $P_e(SNR)$ denote the symbol error probability under the maximum likelihood detection rule.

The next section examines the impact of the randomization matrix constructions on the performance, by quantifying the potential loss in coding gain and diversity gain in comparison to the centralized scheme. Specifically, the *d*iversity order d^* of a scheme with probability of error $P_e(SNR)$ is defined as $d^* = \lim_{SNR\to\infty} -\ln P_e(SNR)/\ln SNR$ and the randomized space-time code is said to achieve a *coding gain* G if $Pe(SNR) \leq G^{-1}SNR^{-d^*}$.

4.3 Performance Analysis

A centralized scheme, with $\mathbf{R} = \mathbf{I}$, under C1) and for $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ has an error rate $O(SNR^{-D(L+1)})$, as well as a coding gain $G = \min_{ij} \prod_{k=1}^{D(L+1)} \alpha_k^{ij}$, where $\alpha_1^{ij} \geq \ldots \geq \alpha_{D(L+1)}^{ij}$ are the eigenvalues of $\mathbf{A}_{ij}^H \mathbf{A}_{ij}$. The simplification in code selection attained through randomization comes at a loss in error rate performance but the loss is modest if the distribution of $\mathbf{r}_t$ is well chosen [82]. The following analysis proves that this is true in a more realistic asynchronous scenario and that additional degrees of diversity can come from the nodes' dispersion. To carry out our analysis we assume that: (**a1**) the symbols have energy E_s, the additive noise is i.i.d.

$\mathcal{CN}(0, N_o)$ and we denote $SNR = E_s/N_o$; (**a2**) each channel impulse response is an independent vector $\mathbf{h}_i \sim \mathcal{CN}(\mathbf{0}, \phi_i \mathbf{\Delta}_h)$, where, for simplicity of analysis, we assume each channel impulse response has the same covariance $\mathbf{\Delta}_h$ scaled by the path loss ϕ_i between the i-th cooperative transmitter and the destination. The overall channel is thus $\mathcal{CN}(\mathbf{0}, \mathbf{\Phi}_h \otimes \mathbf{\Delta}_h)$, where $\mathbf{\Phi}_h$ is a $T \times T$ diagonal matrix containing the nodes' path losses towards the receiver node; (**a3**) the channel vector has an ergodic distribution and packets are sufficiently long such that the sample average performance is close to the corresponding ensemble averages over the possible channel states. One can consider two cases: 1) the nodes keep the randomization coefficients fixed over the transmission of the packet (i.e. $P_e = P_e(SNR|\boldsymbol{\mathcal{R}})$ is a function of a given realization $\boldsymbol{\mathcal{R}}$ of $\mathbf{R}$); 2) the nodes change $\mathbf{r}_i$ at random during the packet so many times that $P_e = E\{P_e(SNR|\boldsymbol{\mathcal{R}})\}$. We study only the first case and state the results for the second case as it is a direct extension of [82].

4.3.1 Diversity Analysis for Static Randomization

Let $\mathbf{Z} = \boldsymbol{\mathcal{R}} \mathbf{\Phi}_h^{1/2} \otimes \mathbf{\Delta}_h^{1/2}$ where the eigenvalue decomposition of $\mathbf{Z}^H \mathbf{Z}$ is $\mathbf{W} \mathbf{\Psi} \mathbf{W}^H$, and $\mathbf{\Psi} = \text{diag}(\psi_1, ..., \psi_{d^*(\mathcal{R})})$ are the ordered nonzero eigenvalues and let $\eta = \text{rank}(\mathbf{\Phi}_h^{H/2} \boldsymbol{\mathcal{R}}^H \boldsymbol{\mathcal{R}} \mathbf{\Phi}_h^{1/2}) \leq \min(D, T)$ and $\eta_L = \text{rank}(\mathbf{\Delta}_h)$. Next, we will equivalently consider $\|(\boldsymbol{\mathcal{R}} \mathbf{\Phi}_h^{1/2} \otimes \mathbf{\Delta}_h^{1/2}) \tilde{\mathbf{h}}\|^2$, $\tilde{\mathbf{h}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ instead of $\|(\boldsymbol{\mathcal{R}} \otimes \mathbf{I})\mathbf{h}\|^2$, $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Phi}_h \otimes \mathbf{\Delta}_h)$. The following theorem provides bounds on the average probability of error for a given realization $\boldsymbol{\mathcal{R}}$:

Theorem 4.3.1. *Let $d^*(\boldsymbol{\mathcal{R}}) = \eta\eta_L$. For a specific $\boldsymbol{\mathcal{R}}$, if C1) is satisfied then:*

$$O(SNR^{-d^*(\mathcal{R})-1}) + \frac{SNR^{-d^*(\mathcal{R})}}{G_L} \leq P_e(SNR|\boldsymbol{\mathcal{R}}) \leq \frac{SNR^{-d^*(\mathcal{R})}}{G_U} \tag{4.3}$$

where

$$G_L = \left(\prod_{k=1}^{d^*(\mathcal{R})} \psi_k\right) \frac{|\mathcal{M}| d^*(\boldsymbol{\mathcal{R}})!}{Q(\max_{ij}(\sqrt{\alpha_1^{i,j}/2}))}$$

$$G_U = \left(\prod_{k=1}^{d^*(\mathcal{R})} \psi_k\right) \left(\min_{ij} \prod_{l=1}^{d^*(\mathcal{R})} \alpha_{D(L+1)-l+1}^{i,j}\right) \frac{4^{-d^*(\mathcal{R})}}{|\mathcal{M}| - 1}$$

Proof. See Appendix C.1 □

The bounds in Theorem 4.3.1 are $O(SNR^{-d^*(\mathcal{R})})$ as $SNR \to \infty$, thus the diversity for a specific realization $\boldsymbol{\mathcal{R}}$ is $d^*(\boldsymbol{\mathcal{R}})$. From the upper bound we can infer the effect of randomization on the coding gain. For comparison we will assume $\boldsymbol{\Phi}_h = \mathbf{I}$, $\boldsymbol{\Delta}_h = \mathbf{I}$, in the centralized scheme $\boldsymbol{\mathcal{R}} = \mathbf{I}$, and for the randomized scheme we assume $E\{\|\mathbf{r}_i\|^2\} = D/T$ for all i. The upper bound shows that the coding gain for the randomized scheme is the coding gain of the centralized scheme scaled by a factor $(\lambda_1 \dots \lambda_\eta)^{(L+1)}(\alpha_1 ... \alpha_{(D-\eta)(L+1)})^{-1}$ where $\eta = \text{rank}(\boldsymbol{\mathcal{R}}^H \boldsymbol{\mathcal{R}})$ and λ_k is the k-th largest eigenvalue of $\boldsymbol{\mathcal{R}}^H \boldsymbol{\mathcal{R}}$. Thus the coding gain loses the contribution of the $(D-\eta)(L+1)$ largest eigenvalues of $\mathbf{A}_{ij}^H \mathbf{A}_{ij}$ in the randomized scheme. The lower bound, however, is only scaled by a factor $(\lambda_1 \dots \lambda_\eta)^{(L+1)}$.

4.3.2 Diversity with Time-Varying Random Coefficients

In the case where the channel and the randomization matrix $\mathbf{R}$ change over the transmission such that the packet experiences multiple realizations, the diversity can be readily obtained extending the results in [82], which considered $P_e(SNR) = E\{P_e(SNR|\boldsymbol{\mathcal{R}})\}$ averaged over all possible outcomes $\boldsymbol{\mathcal{R}}$. The diversity order of the randomized space-time code is always upper bounded by $\eta\eta_L$ where $\eta \leq \min(D, T)$.

Table 4.1: Diversity order: $\eta = \min(T, D)$.

Distribution of $\mathbf{R}$	Condition	Diversity Order
Complex Gaussian	$T = D$	$\eta_L \eta - 1$
Complex Gaussian	$T \neq D$	$\eta_L \eta$
Real Gaussian	$T = D$	$\eta_L \eta - 2$
Real Gaussian	$T = D + 1$ or $T = D - 1$	$\eta_L \eta - 1$
Real Gaussian	$T = D + 2$ or $T = D - 2$	$\eta_L \eta - 0.5$
Real Gaussian	$\lvert T - D \rvert > 2$	$\eta_L \eta$
Uniform Phase	$T = D = 2, \eta_L = 2$	2.5
Real Spherical Distribution	$T = D = 2, \eta_L = 2$	3
Complex Spherical Distribution	$T = D = 2, \eta_L = 2$	4

The following theorem provides sufficient conditions such that the randomized code has full diversity.

Theorem 4.3.2. *Suppose that* $\{\boldsymbol{\mathcal{X}}\}$ *satisfies C1), let* $\mathbf{R}$ *be an* $D \times T$ *have* $p(\mathbf{R})$ *as its probability density function and* $\boldsymbol{\Phi}_h$ *be full rank. For* $D \neq T$*, full diversity* $\min(D, T)\eta_L$ *is achieved if:*

C4) *Rank criterion for* $\mathbf{R}$*: The matrix* $\mathbf{R}$ *is full-rank with probability 1.*

C5) $\mathbb{E}\{|\mathbf{R}^H \mathbf{R}|_{\eta^+}^{-\eta_L}\} < \infty$.

Intuition: If C4) is satisfied, $\eta = \min(D, T)$. Condition C5) is necessary to ensure the product of the η largest singular values of $\mathbf{R}$ is not too small.

In general any randomization matrix $\mathbf{R}$ with independent columns drawn from a continuum distribution satisfies condition C4). Since condition C5) is rather

stringent, a sufficient condition identical to that proposed in [82] is:

Theorem 4.3.3. *Let* $\mathbf{R}$ *be an* $D \times T$ complex *random matrix and denote* $p(\mathbf{R})$ *its probability density function. Assume that the function* $p(\mathbf{R})$ *is bounded and* $tr(\mathbf{R}\mathbf{R}^H) \leq C$ *with probability 1, where* C *is some constant*[2]*. If* $|D-T| > \eta_L$ *and C1) is satisfied, then* $E\{|\mathbf{R}\mathbf{R}^H|_{\eta+}^{-\eta_L}\} < \infty$*. Therefore:*

$$d^* = \begin{cases} T\ \eta_L & \text{if } T < D - \eta_L \\ D\ \eta_L & \text{if } T > D + \eta_L \end{cases} \tag{4.4}$$

Proof. See Appendix C.2. □

The respective diversity for some specific (normalized) randomization schemes are shown in Table 4.1. The distributions are defined as follows: for the complex/real Gaussian distribution, $\{\mathbf{R}\}_{dt} \overset{iid}{\sim} \mathcal{CN}(0,\rho)$; for the Uniform phase distribution $\{\mathbf{R}\}_{di} = a_i e^{j\theta_t[d]}$, $\theta_i[d] \sim U(0, 2\pi)$, while for the real/complex spherical distribution $\mathbf{r}_i = \sqrt{D\rho}\mathbf{g}_i/\|\mathbf{g}_i\|$, $\mathbf{g}_i \overset{iid}{\sim} \mathcal{N}(0,\mathbf{I})$ or $\overset{iid}{\sim} \mathcal{CN}(0,\mathbf{I})$.

We note that if we assume $\boldsymbol{\mathcal{R}}$ is fixed over the transmission due to Jensen's inequality one can prove that $E\{d^*(\boldsymbol{\mathcal{R}})\} \geq d^*$. Interestingly, for large random matrices it can be proven that $d^*(\boldsymbol{\mathcal{R}}) \to E\{d^*(\boldsymbol{\mathcal{R}})\} \to d^*$.

4.3.3 Numerical Results

We simulate the average probability of error for the randomized Time-Reversal Space Time Block Coding (TR-STBC) scheme [59,78] and compare the randomized

[2]In practice, this must be true as the total relay power available to the network is finite

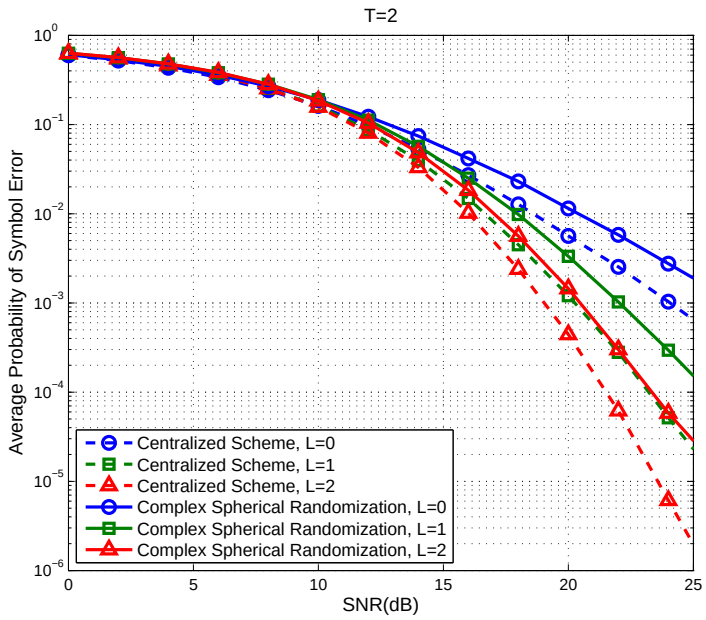

Figure 4.1: Average Probability of Error versus SNR (dB) using TR-STBC scheme where $D = 2$, $T = 2$

and centralized schemes for $D = 2$. In the centralized TR-STBC scheme there are two nodes acting as the first and second antenna respectively. The entries of $\mathbf{R}$ are drawn from a Uniform distribution on a complex hypersphere (with radius $1/\sqrt{T}$) which was observed to perform best among all comparable schemes proposed in [82]. We assume the channels' impulse responses towards the destination are i.i.d. and are due to block fading uncorrelated scattering, i.e. $\mathbf{h}_t \sim \mathcal{CN}(\mathbf{0}, \frac{1}{L+1}\mathbf{I})$, with $\eta_L = L + 1$ where L is the maximum channel memory. We use a block decision feedback equalizer, known to approach ML performance at high SNR.

In all figures we plot the average probability of bit error with respect to $SNR = \frac{E_s}{No}$ using 8PSK. In Fig. 4.1 we examine the case of $T = 2$ nodes for various channel memories in comparison with the centralized scheme. The transmit power for each of the T nodes in both the randomized and centralized schemes is normalized such that $E\{\|\mathbf{r}_i\|^2\} = 1/T$. In Fig. 4.2 we compare the centralized scheme to the

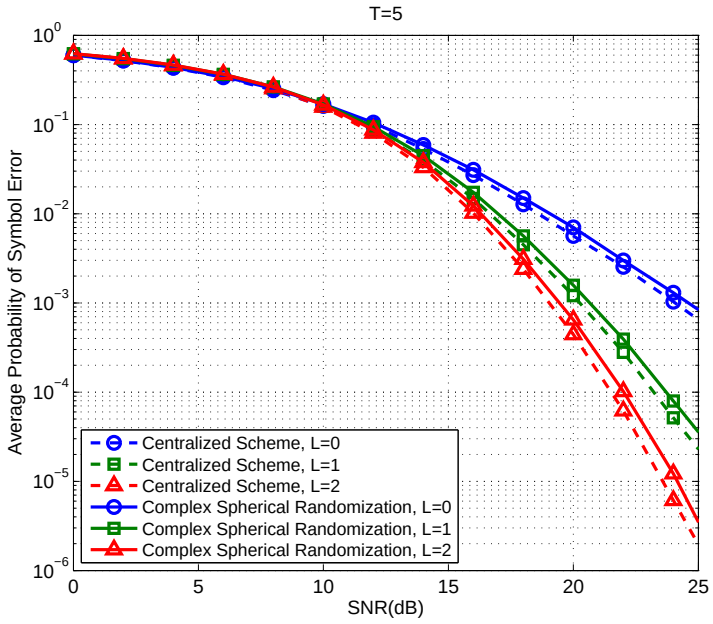

Figure 4.2: Average Probability of Error versus SNR (dB) using TR-STBC scheme where $D = 2$, $T = 5$

case of $T = 5$ nodes for various L. We see as the number of nodes is increased, the performance of the distributed schemes approaches the centralized scheme in both diversity and in coding gain. The theory developed helps describe the trend towards achieving a centralized type performance without requiring much overshoot in the number of needed transmitters; however, diversity and coding gains calculated in theory are nearly impossible to verify at practical ranges of SNRs.

4.4 Cooperative Communication in OTH-HF Channels

We would like to extend the model introduced in Sec. 4.2 to include doubly-selective channels. To accommodate these channels, we modify the $(K + L) \times K$

channel matrix $\mathbf{H}_i$ from the i-th user in (4.1) as:

$$\{\mathbf{H}_i\}_{m,k} = h_i[m, m-k] \tag{4.5}$$

where $h_i[m, m-k]$ is the discrete time time-varying channel impulse response at time l to an input at time $l-k$. Substituting this definition in (4.1), the length $K+L \times 1$ vector of observations:

$$\begin{aligned} \mathbf{y} &= \sum_{i=1}^{T} \mathbf{H}_i \mathbf{G}(\mathbf{s}) \mathbf{r}_i + \mathbf{w} \qquad (4.6) \\ &= \sum_{d=1}^{D} \underbrace{\left(\sum_{i=1}^{T} \mathbf{H}_i r_i(d) \right)}_{\tilde{\mathbf{H}}_d} \mathbf{x}_d + \mathbf{w} \\ &= \sum_{d=1}^{D} \tilde{\mathbf{H}}_d \mathbf{x}_d + \mathbf{w} \qquad (4.7) \end{aligned}$$

where $r_i(d)$ is still the (d,i)-th element of the $D \times T$ randomization matrix $\mathbf{R}$. In the following we provide a brief review of several different transmission/coding schemes. We analyze two single-carrier schemes, that of Space Time Block Coding (STBC) and Time-Reversal Space Time Block Coding (TR-STBC) [59], and a multicarrier transmit diversity scheme known as Space-Frequency OFDM [56].

4.4.1 Space-Time Block Coding (STBC)

The randomized single-carrier modulation scheme for doubly-selective channels can be easily modeled using (4.6), where the columns of the $K \times D$ code matrix $\mathbf{G}(\mathbf{s})$ can be expressed as:

$$\mathbf{x}_d = \mathbf{A}_d \mathbf{s} + \mathbf{B}_d \mathbf{s}^* \tag{4.8}$$

and the matrices $\mathbf{A}_d$ and $\mathbf{B}_d$ uniquely specify the particular space time code.

In order to perform linear or decision feedback equalization of a standard space time code, it is necessary to process several blocks of the received data. If we consider N_b blocks of data, $\mathbf{S} = \text{vec}(\mathbf{s}_0, \mathbf{s}_1, ..., \mathbf{s}_{N_b-1})$, where the channel order L is such that $DL \leq N_b K$, and that guard intervals are placed at the edge of each block, we can express the received signal as:

$$\mathbf{y} = \sum_{d=1}^{D} \tilde{\mathbf{H}}_d \tilde{\mathbf{x}}_d + \hat{\mathbf{w}} \tag{4.9}$$

where the equivalent tall channel $N_b K + L \times N_b K$ matrix $\tilde{\mathbf{H}}_d$ is the same as in (4.7) and the vector $\tilde{\mathbf{x}}_d = (\mathbf{I} \otimes \mathbf{A}_d)\mathbf{S} + (\mathbf{I} \otimes \mathbf{B}_d)\mathbf{S}^*$ is the concatenation of the D columns of $\mathbf{G}(\mathbf{s}_n)$ for $0 \leq n < N_b - 1$. The Kronecker product is denoted by $\otimes$. For purely frequency selective channels, the channel matrices $\tilde{\mathbf{H}}_d$ would have a Toeplitz structure as opposed to the banded structure when the channels are doubly-selective. To perform linear equalization we introduce the vector $\tilde{\mathbf{y}}$

$$\begin{aligned}
\tilde{\mathbf{y}} &= \text{vec}(\mathbf{y}, \mathbf{y}^*) = \mathbf{M}\tilde{\mathbf{s}} + \tilde{\mathbf{w}} \\
\mathbf{M} &= \begin{pmatrix} \sum_{d=1}^{D} \tilde{\mathbf{H}}_l(\mathbf{I} \otimes \mathbf{A}_l) & \sum_{d=1}^{D} \tilde{\mathbf{H}}_d(\mathbf{I} \otimes \mathbf{B}_d) \\ \sum_{d=1}^{D} \tilde{\mathbf{H}}_d^*(\mathbf{I} \otimes \mathbf{B}_d^*) & \sum_{d=1}^{D} \tilde{\mathbf{H}}_d^*(\mathbf{I} \otimes \mathbf{A}_d^*) \end{pmatrix} \\
\tilde{\mathbf{s}} &= \text{vec}(\mathbf{S}, \mathbf{S}^*) \\
\tilde{\mathbf{w}} &= \text{vec}(\hat{\mathbf{w}}, \hat{\mathbf{w}}^*)
\end{aligned} \tag{4.10}$$

where vec$(\cdot)$ denotes the column vector formed by stacking the columns of $(\cdot)$, and $(\cdot)^*$ denotes the complex conjugate. Alternatively we can use the following mapping

$$\begin{aligned}
\tilde{\mathbf{y}} &= \text{vec}(\Re[\mathbf{y}], \Im[\mathbf{y}]) = \mathbf{M}\tilde{\mathbf{s}} + \tilde{\mathbf{w}} \\
\tilde{\mathbf{s}} &= \text{vec}(\Re[\mathbf{S}], \Im[\mathbf{S}])) \\
\tilde{\mathbf{w}} &= \text{vec}(\Re[\hat{\mathbf{w}}], \Im[\hat{\mathbf{w}}])
\end{aligned}$$

Using these models, linear and decision feedback equalizers can be designed.

4.4.2 Time-Reversal STBC

For the Time-Reversal STBC, first proposed in [59], the transmitter splits the data symbols $\mathbf{S}$ into two blocks $\mathbf{S}_1$ and $\mathbf{S}_2$ each of length $N = MN_b/2$ to be transmitted from two antennas. We can then define the vectors $\tilde{\mathbf{x}}_1$ and $\tilde{\mathbf{x}}_2$ in (4.9) as

$$\tilde{\mathbf{x}}_1 = \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{p} \\ -\acute{\mathbf{S}}_2^* \end{bmatrix}, \quad \tilde{\mathbf{x}}_2 = \begin{bmatrix} \mathbf{S}_2 \\ \mathbf{p} \\ \acute{\mathbf{S}}_1^* \end{bmatrix} \tag{4.11}$$

where the vector $\mathbf{p}$ consists of zero guard intervals or a training sequence in order to avoid inter-block interference (IBI), and $\acute{(\cdot)}$ denotes the time-reversal operation. At the receiver, the second received block, $\acute{\mathbf{y}}_2^*$, is time reversed and conjugated, then combined with the first received block $\mathbf{y}_1$:

$$\underbrace{\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} \tilde{\mathbf{H}}_1 & \tilde{\mathbf{H}}_2 \\ \acute{\mathbf{H}}_2^* & -\acute{\mathbf{H}}_1^* \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} \mathbf{S}_1 \\ \mathbf{S}_2 \end{bmatrix}}_{\mathbf{S}} + \underbrace{\begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix}}_{\mathbf{w}} \tag{4.12}$$

After processing the received signal vector $\mathbf{y}$ with the spatio-temporal matched filter $\mathbf{H}^H$ we have

$$\tilde{\mathbf{y}} = \mathbf{H}^H\mathbf{y} = \mathbf{H}^H\mathbf{H}\mathbf{S} + \mathbf{H}^H\mathbf{w} \tag{4.13}$$

In this case, with frequency-selective channels constant over the two blocks, the spatio-temporal matched filter perfectly decouples the decoding of the data blocks $\mathbf{S}_1$ and $\mathbf{S}_2$. However, in the case of doubly-selective channels, the structure of the channel matrices is no longer Toeplitz but banded. As shown in [38], the

performance of the TR-STBC scheme is much worse in the case of doubly-selective channels and two extensions were proposed for the scheme: 1) shortening the data block length and 2) extending the matched filtering to the time-varying channel.

4.4.3 Space-Frequency OFDM

In the SF-OFDM scheme the data symbols $\mathbf{S}$ are coded into a $N \times D$ matrix $\tilde{\mathbf{X}}$ according to code matrix $\mathbf{G}(\mathbf{s})$:

$$\underbrace{\begin{bmatrix} \tilde{\mathbf{x}}_1 & \tilde{\mathbf{x}}_2 & \dots & \tilde{\mathbf{x}}_D \end{bmatrix}}_{\tilde{\mathbf{x}}} = \begin{bmatrix} \mathbf{G}(\mathbf{s}_0) \\ \mathbf{G}(\mathbf{s}_1) \\ \vdots \\ \mathbf{G}(\mathbf{s}_{Q-1}) \end{bmatrix} \tag{4.14}$$

Each of the columns of $\tilde{\mathbf{X}}$ is then modulated by an N-point inverse discrete Fourier transform (IDFT) and a cyclic prefix N_{cp} ($L \leq N_{cp} \leq N$) is added before transmitting over a frequency-selective channel of order L. Assuming the channel impulse responses remain constant over the entire block interval, the received signal vector $\mathbf{y}$ can be expressed as:

$$\mathbf{y} = \sum_{d=1}^{D} \tilde{\mathbf{H}}_d \mathbf{F}^H \tilde{\mathbf{x}}_d + \mathbf{w} \tag{4.15}$$

where $\mathbf{F}^H$ denotes the IDFT operation and addition of the N_{cp} length cyclic prefix. For the frequency selective case, the convolutions are cyclic, therefore at the receiver, the multi-path corrupted cyclic prefix is removed and receiver computes the N-point DFT of $\mathbf{y}$. Since the DFT of a cyclic convolution in the time-domain results in multiplication in the frequency domain, the demodulated signal vector

$\mathbf{Y}$ is given by:

$$\mathbf{Y} = \sum_{d=1}^{D} \mathbf{\Lambda}_d \tilde{\mathbf{x}}_d + \tilde{\mathbf{w}} \tag{4.16}$$

where $\mathbf{\Lambda}_d$ is a diagonal subcarrier decoupling matrix whose elements are the DFT's of the respective frequency-selective channel impulse responses $\mathbf{h}_d$. For the case where the code matrix corresponds to the Alamouti space time code [5], the space-frequency decoder can construct the decision estimate vector $\hat{\mathbf{S}}$ as:

$$\begin{aligned} \hat{\mathbf{S}}_e &= \frac{\mathbf{\Lambda}_{1,o}^* \mathbf{Y}_e + \mathbf{\Lambda}_{2,e} \mathbf{Y}_o^*}{(\mathbf{\Lambda}_{1,o}^* \mathbf{\Lambda}_{1,e} + \mathbf{\Lambda}_{2,o}^* \mathbf{\Lambda}_{2,e})} \\ \hat{\mathbf{S}}_o &= \frac{-\mathbf{\Lambda}_{1,e} \mathbf{Y}_o^* + \mathbf{\Lambda}_{2,0}^* \mathbf{Y}_e}{(\mathbf{\Lambda}_{1,o}^* \mathbf{\Lambda}_{1,e} + \mathbf{\Lambda}_{2,o}^* \mathbf{\Lambda}_{2,e})} \end{aligned} \tag{4.17}$$

where $(\cdot)_e$ and $(\cdot)_o$ correspond to the vectors created from the even and odd components of $(\cdot)$ respectively. In the case of a matrix, $(\cdot)_e$ and $(\cdot)_o$ correspond to the matrices whose diagonal consists of the even and odd elements of the diagonal of $(\cdot)$.

For the case of doubly-selective channels, the matrices $\mathbf{\Lambda}_d$ are no longer diagonal but now have entries on the sub and super diagonals. These banded matrices introduce ICI, complicating the symbol estimation task.

4.4.4 Modeling the HF Channel

In order to establish long distance communication (over the horizon) we make use of the HF channel. We model the doubly-selective channel using the Watterson channel model [97]. The model consists of a tapped-delay line modeling the delays τ (on the order of milliseconds) of the multi-path time-varying environment. Specifically, each path is modeled by a time-varying tap-gain function $c_q(t)$ characterized by its doppler spread σ_{dq}^2 and doppler shift f_{dq}, i.e. $h(t,\tau) = \sum_{q=1}^{Q} c_q(t)\delta(\tau - \tau_q)$.

As verified by Watterson, each tap-gain function can be modeled by an independent zero-mean complex-Gaussian stationary random process for each of the two resolvable magnetoionic components, whose power spectral density is the following Gaussian function:

$$\begin{aligned} \nu_q(f) &= \frac{C_{qa}(0)}{\sqrt{2\pi\sigma_{dqa}^2}} \exp\left[-\frac{(f-f_{dqa})^2}{2\sigma_{dqa}^2}\right] \\ &+ \frac{C_{qb}(0)}{\sqrt{2\pi\sigma_{dqb}^2}} \exp\left[-\frac{(f-f_{dqb})^2}{2\sigma_{dqb}^2}\right] \end{aligned} \tag{4.18}$$

where the subscripts a and b correspond to the two resolvable magnetoionic components and the subscript q corresponds to the path number. In our analysis, we use only one Gaussian component for $c_q(t)$ with negligible doppler shift ($f_{dq} = 0$) as recommended in [3]. The ratio of the output power to the input power is then specified by $C_v(0)$. The ith discrete time channel impulse response is then given by:

$$h_i[m,k] = \sum_{q=1}^{Q} c_{q,i}[m]p_{q,i}[k] \tag{4.19}$$

where $p_{qi}[k] = p(kT_s - \tau_{qi})$ is a Nyquist pulse shape with bandwidth B, and $c_{q,i}[m] = c_{qi}(mT_s)$ is the qth discrete-time baseband tap-gain function corresponding to delay $\tau_{q,i}$ at time mT_s.

4.4.5 Performance Comparison of Coding Schemes

The system parameters are chosen according to current military standards [2]. Specifically, we consider alternating blocks of pilot and data symbols, with a bit-rate of 4.8 kbits/sec, and uncoded 8PSK symbol transmission. The lengths of the probe and data blocks are 16 and 32 symbols respectively. In all simulations the number of virtual antenna $D = 2$.

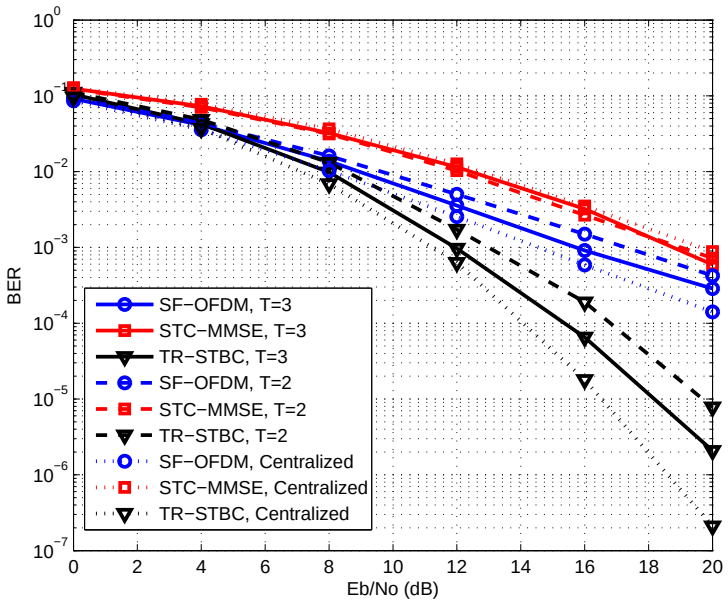

Figure 4.3: Randomized Cooperative Scheme for T cooperating nodes for a frequency selective channel with order $L = 2$.

We first consider the randomized cooperative scheme in a purely frequency selective channel with varying number of cooperative nodes. We observed in simulation that complex spherical random vectors had the best performance among the distributions examined for $\mathbf{R}$, hence we let the i-th column of the matrix $\mathbf{R}$ be selected uniformly on the surface of a complex hypersphere of radius $\|\mathbf{r}_i\| = \sqrt{2/T}$. We compare the obtainable diversity of the randomized cooperative scheme compared to the centralized scheme with two transmit antenna and one receive antenna for a channel of order $L = 2$, perfectly estimated at the receiver in Fig. 4.3

In Fig. 4.4 we show the randomized cooperative scheme for T nodes in the 'poor' channel model as defined in [1], equivalent to 'moderate' conditions in low-latitudes in [3]. The differential time delay $\tau = 2$ ms and doppler spread $\sigma_d^2 = 1.5$ Hz for a two component multi-path fading environment with equal mean attenuation. We assume the channels seen by each antenna are independent with equiva-

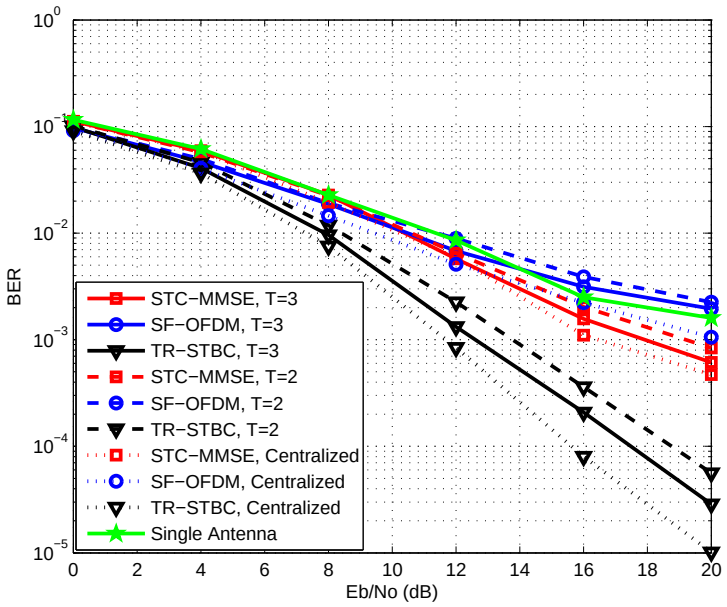

Figure 4.4: Randomized Cooperative Scheme for T cooperating nodes for the 'poor' channel model.

lent parameters. At the receiver, we assume the channel is constant over the block and a perfect channel estimate has been obtained at the beginning of each block. We also show the single antenna scheme for comparison.

In Fig. 4.5 we examine the BER of the randomized cooperative scheme again with T nodes and a fixed $SNR = E_b/No = 15$ dB, differential time delay $\tau = 0.5$ ms for varying doppler spreads and equal mean attenuation of a two component multi-path channel. Fig. 4.6 shows the same setup but for differential time delay $\tau = 2$ms.

As can be seen in Fig. 4.3, as the number of cooperating nodes T increases, the diversity approaches the diversity of the centralized scheme with $L = 2$. The TR-STBC and space-time coded schemes with an MMSE equalizer (STC-MMSE) obtain diversity through both space and time with uncoded transmission. As SNR

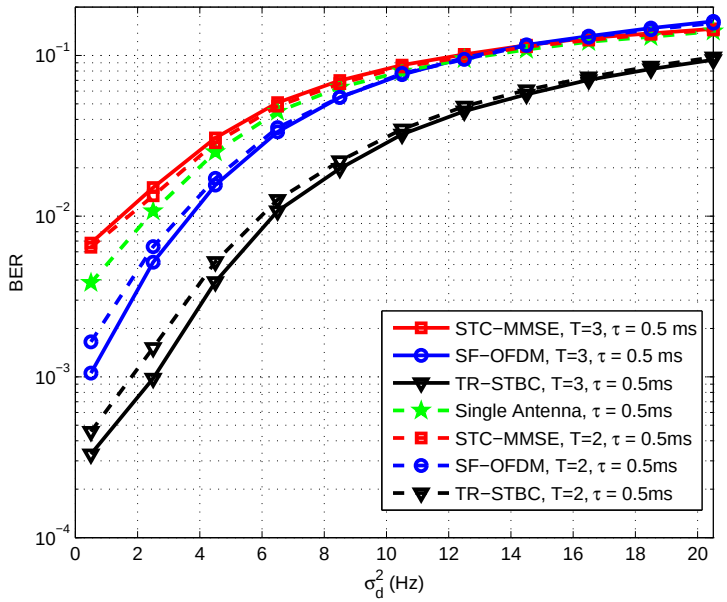

Figure 4.5: Randomized Cooperative Scheme for T cooperating nodes for varying doppler spread and differential time delay $\tau = 0.5$ ms.

becomes large enough, both schemes will outperform the SF-OFDM scheme which obtains only spatial diversity.

In Fig. 4.4, we see that the introduction of the HF channel decreases the performance of the schemes, and the BER starts to exhibit an error floor at higher SNR. The TR-STBC scheme significantly outperforms the other schemes for the 'poor' HF channel model. At a bit error rate (BER) of 10^{-3} the TR-STBC scheme provides about 5 dB of gain over the STC-MMSE scheme. By implementing the extensions proposed in [38], the gain could be further increased.

The schemes all deteriorate rapidly as the doppler spread increases, as shown in Fig. 4.5. Again the TR-STBC scheme has the best performance among the schemes for smaller values of doppler spread. In Fig. 4.6, it is shown that the TR-STBC and STC-MMSE schemes have increased BER for larger values of delay

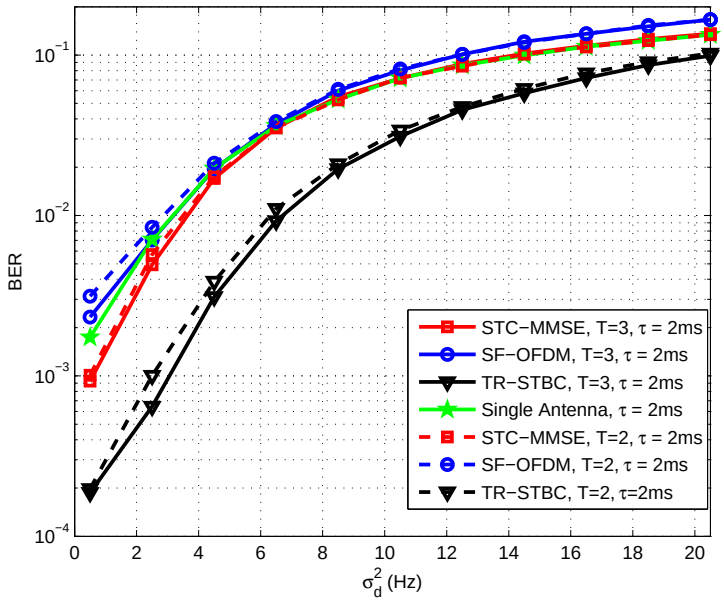

Figure 4.6: Randomized Cooperative Scheme for T cooperating nodes for varying doppler spread and differential time delay $\tau = 2$ ms.

spread, while the performance of the SF-OFDM scheme is inversely related to the delay spread, as mentioned previously.

The potential power gain from multiple cooperative nodes is omitted from these curves, though this is an obvious benefit of this architecture. Each additional node can potentially produce additional power gains assuming the path attenuations are of the same order. However, a shadowing model for the large scale fading information is not available in a form that can be easily simulated.

4.5 Summary

In this chapter we analyzed the diversity obtainable for the randomized cooperative communication scheme proposed in [82] for frequency-selective channels. It was

shown that this scheme can reap diversity not only from the coding scheme used but also through the frequency-selectivity of the channel. Through simulation it was observed that as the number of cooperating radios is increased, even moderately, the performance of the distributed cooperative communication scheme approaches that of the centralized scheme, both in diversity and coding gain.

The application of this cooperative communication scheme to HF channels was then discussed, and through simulation we compared the performance of several different randomized coding schemes in the presence of an HF channel, assuming that the radios clocks were synchronized so that no carrier offsets were present in the system. It was observed that the TR-STBC had the best performance among the coding schemes examined.

We conclude that effects of dispersion in time do not appear to be problematic as long as the channel can be reliably estimated. In practice, the effects of Doppler spreading and shifting reduce the effectiveness of code designs that are aimed only at block-fading channels. In the next chapter we will utilize the ideas introduced in Chapter's 2 and 3 to address the channel estimation problem and study numerically the system performance when the CSI is estimated using these techniques.

CHAPTER 5

SPARSE CHANNEL ESTIMATION FOR ASYNCHRONOUS COOPERATIVE COMMUNICATION IN OTH-HF CHANNELS

5.1 Motivation

In the preceding chapter, we made the assumption that the time-variations in the cooperative channel were due to the physical propagation environment. We simulated the performance of this scheme using the Watterson model [97] for HF channels, but neglected the affects of the carrier offsets resulting from the asynchrony among the nodes. The 'virtual' time-varying channel impulse response corresponding to the d-th column of the code was assumed to be:

$$\tilde{h}_d(t,\tau) = \sum_{i=1}^{T} r_i(d) \sum_{q=1}^{Q} c_{q,i}(t)\delta(\tau - \tau_{q,i})$$

where $r_i(d)$ was the randomization coefficient weighting the d-th column of the code for the i-th cooperative radio, $c_{q,i}(t)$ was the time-varying tap-gain function corresponding to the q-th propagation path of the i-th user, defined using the Watterson model, and $\tau_{q,i}$ its delay. However, assuming non-negligible carrier offset ξ_i between the i-th node and the receiver, the channel now is now modeled as:

$$\tilde{h}_d(t,\tau) = \sum_{i=1}^{T} r_i(d) e^{j2\pi\xi_i t} \sum_{q=1}^{Q} c_{q,i}(t)\delta(\tau - \tau_{q,i})$$

The addition of these carrier offsets complicates the estimation of the discrete-time baseband equivalent channel impulse response. We note that the time-variations on each of the D 'virtual' channels are caused by the carrier offsets, which are common to all the virtual channels, as well as the time-variations resulting from the Watterson channel. Due to the presence of multiple carrier offsets,

the independent time-varying tap-gain functions, and the presence of a cooperative channel, common carrier/phase tracking methods such as the phase-locked loop and Costas loop are unable to track and compensate for these offsets. These offsets must then be accounted for in the channel estimation stage of the receiver. In this chapter we apply the methods discussed in Chapters 2 and 3 to these doubly-selective cooperative channels.

5.2 Randomized Cooperative Communication in the Presence of Carrier Offests

For the moment, let us assume that the Doppler spreading due to the ionospheric propagation is negligible in comparison with the carrier offsets introduced due to asynchrony. Denoting $s_i(t) = \sum_{d=1}^{D} \sum_{k=0}^{K-1} x_d[k] r_i(d) p(t - kT_s)$ as the signal transmitted from the i-th cooperative radio, the observations:

$$y(mT_s) \;=\; \sum_{i=1}^{T} \sum_{q} s_i(mT_s - \tau_{q,i}) e^{j2\pi\xi_i mT_s} c_{q,i} \tag{5.1}$$

Different from the MISO system considered in Sec. 2.2, we make no assumption about the cooperative radios forming a uniform linear array, and further, do not assume the set of propagation delays between each cooperative radio and the receiver. Note also that the time-variations in this model are due to carrier offsets, and thus vary only between cooperative radios, not between the Q propagations paths between each cooperative radio and the receiver.

After expanding the term $s_i(mT_s - \tau_{q,i})$ we have the following equivalent model:

$$y(mT_s) = \sum_{d=1}^{D} \sum_{k=0}^{K-1} \tilde{h}_d[m, m-k] x_d[k] + w[k] \tag{5.2}$$

where the d-th virtual channel:

$$\begin{aligned}\tilde{h}_d[m, m-k] &= \sum_{i=1}^{T} e^{j2\pi\xi_i m T_s} r_i(d) \sum_q p((m-k)T_s - \tau_{q,i}) c_{q,i} \\ &= \sum_{v=1}^{V=TQ} e^{j2\pi\xi_v m T_s} p((m-k)T_s - \tau_v) a_{v,d} \end{aligned} \tag{5.3}$$

where each pair (ξ_v, τ_v) corresponds to a pair $(\xi_i, \tau_{q,i})$ and the likewise the coefficient $a_{v,d}$ corresponds to some $c_{q,i} r_i(d)$.

Similar to the vector channel model in (4.2), assuming the channel has finite length L:

$$\begin{aligned}\mathbf{y} &= \sum_{d=1}^{D} \boldsymbol{\mathcal{X}}_d \mathbf{h}_d + \mathbf{w} = \begin{bmatrix} \boldsymbol{\mathcal{X}}_1 & \dots & \boldsymbol{\mathcal{X}}_D \end{bmatrix} \begin{bmatrix} \mathbf{h}_1 \\ \vdots \\ \mathbf{h}_D \end{bmatrix} + \mathbf{w} \\ &= \boldsymbol{\mathcal{X}} \mathbf{h} + \mathbf{w} \end{aligned} \tag{5.4}$$

where $\boldsymbol{\mathcal{X}}$ and $\mathbf{h}$ are defined similarly to (3.3). Using Prop. 2.2.1, and assuming the same channel basis $\boldsymbol{G}$ for each of the channels $\mathbf{h}_1, ..., \mathbf{h}_D$:

$$\mathbf{y} = \begin{bmatrix} \boldsymbol{\mathcal{X}}_1 & \dots & \boldsymbol{\mathcal{X}}_D \end{bmatrix} (\mathbf{I}_D \otimes \boldsymbol{G}) \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \vdots \\ \boldsymbol{\alpha}_D \end{bmatrix} + \mathbf{w} \tag{5.5}$$

In Sec.'s 2.4.1 and 3.5 we provided a SI input for the SISO doubly-selective channel that consisted of a linear chirp, modified it using the metric of localized coherence, and approximated this design with a linear modulated signal in Sec. 3.5. One possible way to use this design for the current model (5.5), is to split the training such that any one time, only one of the virtual channels contains training, while

the others are silent:

$$\mathbf{y} = \begin{bmatrix} \boldsymbol{\mathcal{X}}_1\boldsymbol{G} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\mathcal{X}}_2\boldsymbol{G} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \dots & \ddots & \boldsymbol{\mathcal{X}}_D\boldsymbol{G} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_D \end{bmatrix} + \mathbf{w} \tag{5.6}$$

Then, assuming the training corresponding to each of the $\boldsymbol{\mathcal{X}}_d$ was sufficiently informative, the set of training inputs over the entire cooperative system would be sufficiently informative. However, assuming the amount of resources devoted to training was limited, this is not the best training design for this cooperative system. As will be demonstrated via localized coherence and verified via simulation, transmitting training concurrently on each of the virtual channels results in both better channel MSE performance, as well as overall bit-error rate performance. This will be demonstrated using the TR-STBC scheme discussed in Sec. 4.4.2 with $D = 2$ virtual channels, where on one channel we use the down-up chirp training design on one channel, and an up-down chirp on the other channel.

5.3 Improving Channel Estimation Performance

We assume that each packet transmitted in our cooperative schemes contains several blocks, each block corresponding to a training sub-block followed by a data sub-block. We initially assumed that the time-variations on each of the channels was a function solely of the carrier-offset, and not on the propagation environment. When including the Watterson HF channel in the model, we may assume that the Doppler spread is negligible in comparison to the carrier offsets. We note that for sufficiently small block sizes, we may assume that the coefficients $\boldsymbol{\alpha}$ change slowly over the transmission of several blocks. Further the model (5.3) can poten-

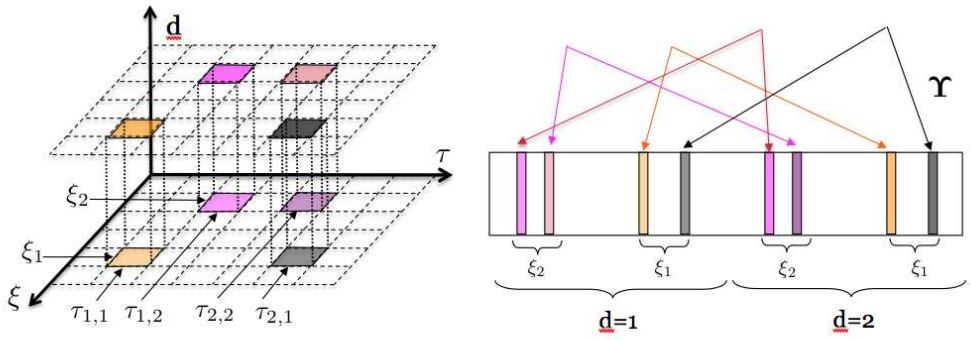

Figure 5.1: Example channel support structure for cooperative channel.

tially capture the contribution from the Doppler spreading of the HF propagation environment.

It is easy to demonstrate that for appropriately designed training signals, the localized coherence is decreased when increasing the number of observations, due to the potential reduction in the inter-column correlation between columns of $\boldsymbol{\mathcal{X}}\boldsymbol{G}$. Taking advantage of the packet structure of our system, when estimating the channel, we can combine the training observations over several blocks, decreasing correlation between columns of the resulting sensing matrix. However, including too many blocks in the estimation of the channel can cause a decrease in performance as a result of the time=variations in the coefficients $\boldsymbol{\alpha}$ due to the Doppler spreading.

Estimation of the channel can further be improved by noting that the support of each of the vectors $\boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_D$ is the same. This is due to the fact that each of the virtual channels is a random linear combination of the T channels between the cooperative nodes and the receiver. In (5.3), we observe that each each $a_{v,d}$ corresponds to a particular pair of parameters (ξ_v, τ_v), and under the assumptions of Prop. 2.2.1, each $\boldsymbol{\alpha}_1, .., \boldsymbol{\alpha}_D$ will have the same support as demonstrated in Fig.

5.1. To capitalize on this structure, when searching for a sparse solution, we can constrain the set of solutions to only those such that $\boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_D$ have the same support. This is easily implemented using as a starting point, the OMP [69] or MP algorithm's [64]. In these algorithm, at each iteration, the component of the sensing matrix most correlated with the residual signal is selected, building a representation dictionary for the observations $\mathbf{y}$. Modifying this algorithm, so that at each iteration we select one component from each matrix $\boldsymbol{\mathcal{X}}_d\boldsymbol{G}$, while constraining the index of this component to be the same in each matrix. The component selected in each basis is chosen as the one maximizing $\max_i \sum_{d=1}^{D} |(\boldsymbol{\mathcal{X}}_d\boldsymbol{G})_i^H \mathbf{r}^{(n)}|$ where $\mathbf{r}^{(n)}$ is the residual signal at the n-th iteration of the algorithm and $(\boldsymbol{\mathcal{X}}_d\boldsymbol{G})_i$ is the i-th column of the sensing matrix corresponding to the d-th virtual channel. A similar method known as Multivariate Matching Pursuit (MMP) algorithm was described in [31] [42]. However, in MMP multiple we have available a mixture of multiple observation vectors, each corresponding to a single signal with the same support over a single dictionary. Different from MMP, though we observe a mixture of multiple signals, each corresponding to a virtual channel, each of our signals has a different dictionary (corresponding to $\boldsymbol{\mathcal{X}}_d\boldsymbol{G}$), but is assumed to have the same support over each of these dictionaries. As will be shown in the following numerical results section, when constraining the support in this manner, a slight improvement can be obtained in the channel estimation performance.

5.4 Numerical Results

In the following, we will use the orthogonal matching pursuit algorithm (OMP) described in [69], as well as the modified OMP algorithm described above, constraining the estimate of the support to be the same over each $\boldsymbol{\mathcal{X}}_d\boldsymbol{G}$, which we

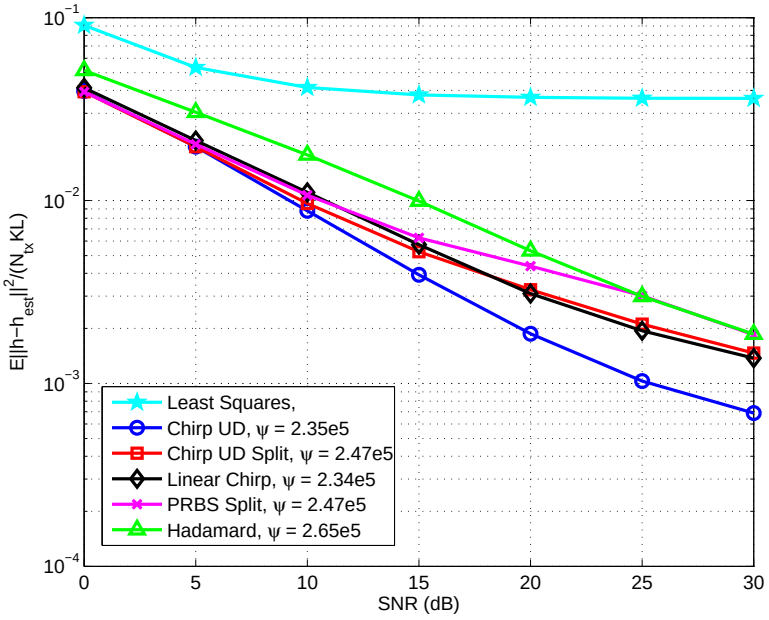

Figure 5.2: Channel MSE versus SNR (dB) for Asynchronous Cooperative HF Channel with $T = 3$ users and $D = 2$ virtual channels.

refer to as the Support Constrained OMP algorithm (SCOMP).

In all simulations, we implement the randomized cooperative communication scheme using TR-STBC as described in Sec. 4.4.2 with $D = 2$ virtual antennas. The system parameters are chosen according to current military standards [2]. Specifically, we consider alternating blocks of pilot and data symbols, with a signal bandwidth of 2.4 kHz, and using uncoded QPSK symbol transmission. The pilot and data blocks are each 37 symbols long, unless otherwise specifed. Each column of the randomization matrix $\mathbf{R}$ is selected uniformly on the surface of a complex hypersphere of radius $\|\mathbf{r}_i\| = \sqrt{1/(DT)}$, such that the average received power is unity regardless of the size of the code D or number of cooperative radios T.

In the simulations, the channel between each cooperative radio and the receiver is generated using the 'poor' channel as defined in [1], or equivalently the 'moderate'

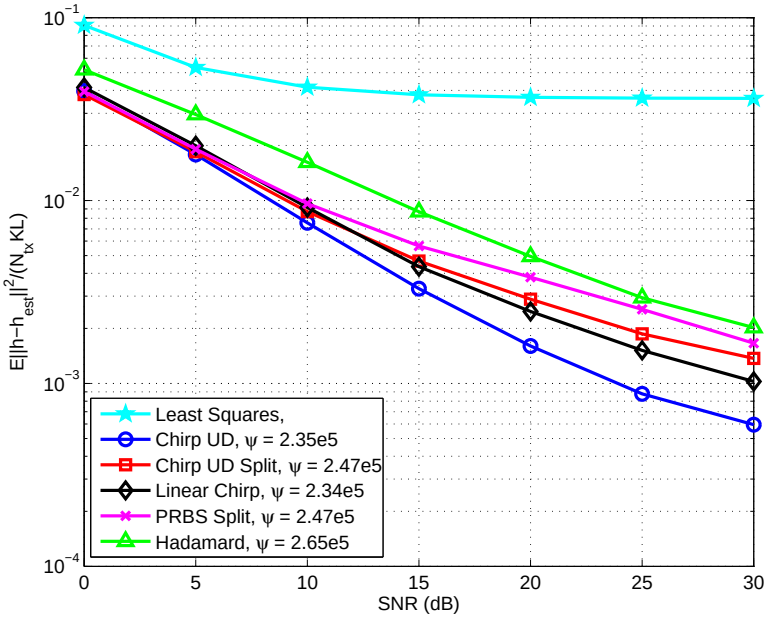

Figure 5.3: Channel MSE versus SNR (dB) for Asynchronous Cooperative HF Channel with $T = 3$ users and $D = 2$ virtual channels, using SCOMP.

channel in low-latitudes as specified in [3]. The differential time delay on each of these channels (due to the propagation environment as well as possible asynchrony among the cooperative radios) is generated uniformly at random $\tau_{q,i} \in (0, 2)$ ms and the doppler spread $\sigma_d^2 = 2$ Hz for a two component HF multi-path fading environment with equal mean attenuation is generated according to the Watterson HF model [97]. Additionally, due to the assumption of asynchrony among nodes, we add a uniformly distributed carrier offset $\xi_i \in (-\xi_{max}/2, \xi_{max}/2)$ the channel between each cooperative radio and the receiver.

In Fig.'s 5.2, 5.3 we plot the simulated MSE of the channel estimate for $T = 3$ cooperative radios using both OMP and SCOMP for various training designs and include their localized coherence values ψ, while in Fig. 5.4 we plot the simulated MSE of the predicted channel estimate over the subsequent data block. All MSE's

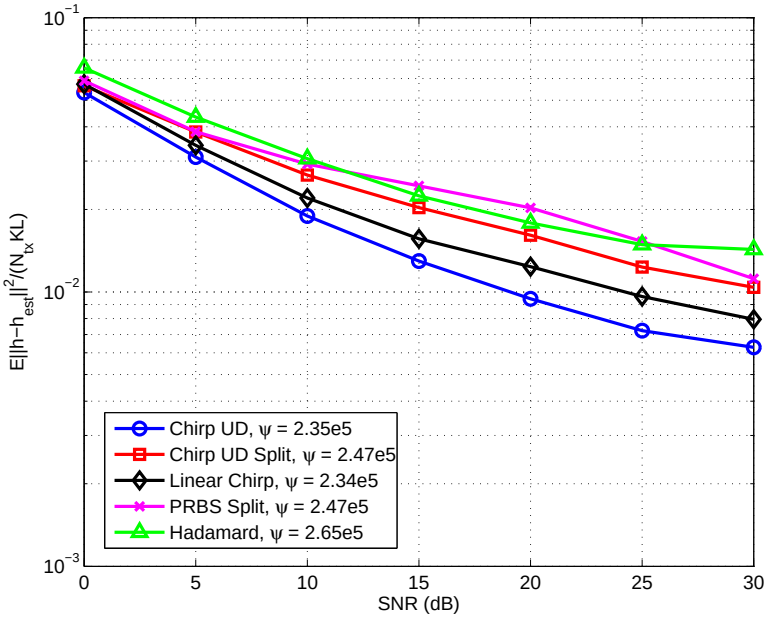

Figure 5.4: Predicted channel MSE versus SNR (dB) for Asynchronous Cooperative HF Channel with $T = 3$ users and $D = 2$ virtual channels, using SCOMP.

are normalized by the number of equivalent taps in the channel L, the number of observations of the channel K, and the number of virtual channels D. The training sequences are each $N_t = 30$ symbols long.

The first several training designs are based on a linearly modulated version of the down-up chirp described in Sec. 3.5. In the first design, on the first virtual channel we transmit an up-down chirp where the training symbols $x_1[n] = e^{j2\pi\beta n^2/N_t}$ for $n = 0, ..., 14$ and $e^{-j2\pi\beta n^2/N_t}$ for $n = 15, ..., 29$, while the second virtual channel transmits $x_2[n] = x_1^*[n]$ corresponding to a down-up chirp. The chirp rate is set as $\beta = 2(1 - \xi_{max})$ and we denote this design as Chirp UD in the figures. The second design, denoted as CHIRP UD Split, separates the training on each virtual channel in time, as in (5.6), where on the first virtual channel we again transmit a up-down chirp, however only 15 symbols long, in the first 15 sym-

bols, followed by zeros, while on the second channel the first 15 symbols are zero, while the remaining 15 symbols correspond to a down-up chirp. The third design, denoted as Linear Chirp simply transmits a linear chirp with increasing frequency on the first virtual channel $x_1[n] = e^{j\pi\beta n^2/N_t}$, while simultaneously transmitting a linear chirp with decreasing frequency on the second virtual channel $x_2[n] = x_1^*[n]$.

The final two training designs are based on binary sequences. The first design, called $\text{PRBS} - \text{Split}$, transmits a length 15 pseudo-random binary sequence on virtual channel one, while simultaneously transmitting zeros on the second virtual channel. In the second half of the training block, the second virtual channel now transmits the pseudo-random binary sequence, while no training is present on the first virtual channel. The final training design, denoted as Hadamard, transmits on each virtual channel two different length 32 Hadamard sequences, shortened to a length of 30 symbols, on each of the virtual antenna.

We observe that as predicted by the localized coherence, the designs that separate the training on the virtual channels in time, as in (5.6), do not perform as well as the training designs that simultaneously transmit training on both virtual channels. Further, the training design that transmits an up-down chirp on one channel, while simultaneously transmitting a down-up chirp on the other virtual channel outperforms the design that simply transmits an up-chirp on the first virtual channel and a down-chirp on the second virtual channel. Further, a slight improvement in the estimation performance is gained when using the SCOMP algorithm instead of the OMP algorithm to estimate the channels.

In order to improve the estimation performance, the estimate of the predicted channel is obtained using the the observations from the training blocks on either side of a particular data block. Once an estimate of $\boldsymbol{\alpha}$ is obtained, we predict the

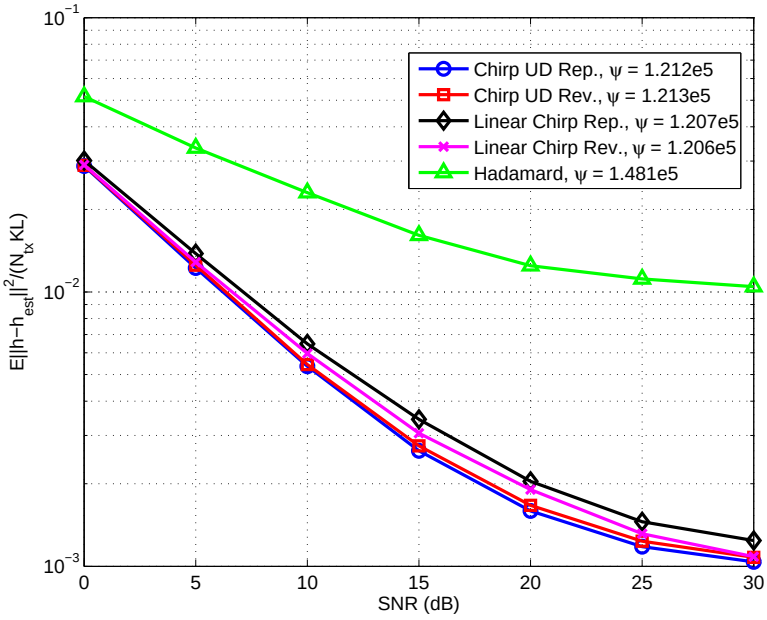

Figure 5.5: Predicted channel MSE versus SNR (dB) for Asynchronous Cooperative HF Channel with $T = 3$ users and $D = 2$ virtual channels, using SCOMP and multiple training blocks.

channel using the model (5.3) over the samples K corresponding to the data. In Fig. 5.5, we plot the predicted channel MSE when using two training blocks, the ones immediately preceding and succeeding the current data block. As expected, increasing the number of observations for the training sequence results in a decrease of the localized coherence. The Chirp UD, Linear Chirp, and Hadamard designs are the same as above, however for the Chirp UD and Linear Chirp designs, we compare the case of repeating the training in each training block (denoted as Chirp UD Rep. and Linear Chirp Rep.) and flipping the training on the virtual channels in each alternating block (denotes as Chirp UD Rev. and Linear Chirp Rev.). For the Hadamard design, we choose two different Hadamard sequences in each successive training block, again shortening the sequences to 30 symbols. We observe, as expected, that increasing the number of observations

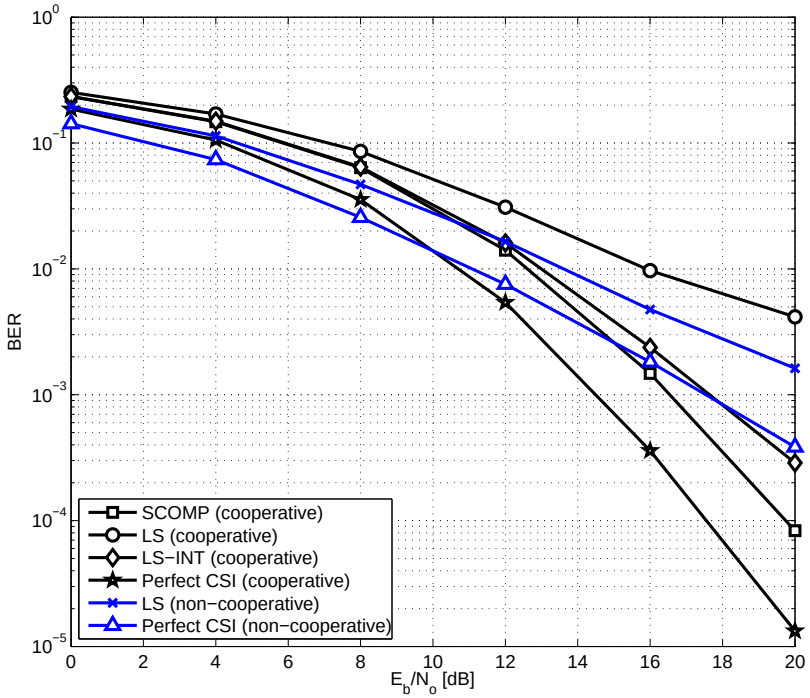

Figure 5.6: Average Probability of Error versus SNR (dB) using TR-STBC $T = 3$ users and $D = 2$ virtual channels. Assumes asynchronous cooperative HF Channel with maximum carrier offset $\xi_{max} = 10$ Hz.

used for estimating the channel results in better localized coherence, as well as estimation performance. Further, it appears that the Chirp UD design offers slightly improved performance compared to the Linear Chirp, while the design based on Hadamard sequences is less desirable in terms of estimation performance, as predicted by the localized coherence. The alternating of the training sequences in successive training blocks does not appear offer any improvement in the channel estimation performance for these designs.

In Fig.'s 5.6, 5.7, 5.8 we simulate the bit error rate (BER) performance of the cooperative system described above for $\xi_{max} = 10, 20, 30$ respectively. Based on the simulated predicted channel estimation MSE, we use the Chirp UD training design

for each of these simulations repeating the training in each successive training block. A block decision feedback equalizer is then used on these observations. For the cooperative system, we also include the BER performance assuming perfect channel state information (CSI), as well as the BER performance when using least squares channel estimation, as well as a channel prediction method that linear interpolates the channel estimates obtained via least squares on the two training blocks surrounding a data block. As a comparison, we include a SISO link (denoted as non-cooperative in the simulations) using both perfect CSI and a least squares channel estimate, assuming the carrier offset has been accounted for in the front-end of the receiver using a PLL.

We observe that the SCOMP channel estimation method outperforms the least squares channel estimation methods in the cooperative channel. As the carrier offset ξ_{max} is increased, the least-squares channel estimation methods are not able to cope with the increased time-variations in the cooperative channel. For smaller values of carrier offset ($\xi_{max} = 10, 20$ Hz), the cooperative transmission scheme using the SCOMP channel estimator is able to harvest some diversity out of the system. In fact, for the case of $\xi_{max} = 10$ Hz, the system achieves the same diversity as the perfect CSI case, as demonstrated in Fig. 5.6. As the carrier offsets increase, the diversity gain starts to diminish, and in the case of $\xi_{max} = 30Hz$ in Fig. 5.8 we no longer have a diversity gain.

In comparison with the non-cooperative SISO system, which assumes the carrier offsets has already been compensated for, the cooperative scheme outperforms the non-cooperative SISO system at moderate values of SNR as demonstrated in Fig.'s 5.6 and 5.7. However, when the carrier offsets become too large, the SISO non-cooperative system has better BER performance, as observed in Fig. 5.8.

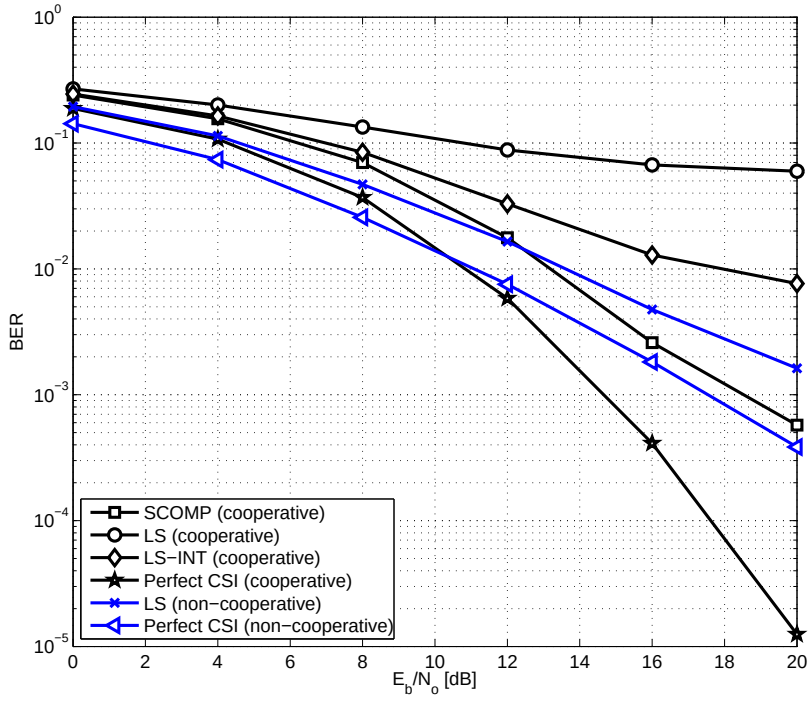

Figure 5.7: Average Probability of Error versus SNR (dB) using TR-STBC $T = 3$ users and $D = 2$ virtual channels. Assumes asynchronous cooperative HF Channel with maximum carrier offset $\xi_{max} = 20$ Hz.

5.5 Summary

In this chapter, we introduced the effects of asynchrony into the cooperative system, where each cooperative radio had a carrier offset relative to the receiver, greatly increasing the time-variations in the equivalent channel observed at the receiver. In order to estimate this doubly-selective channel, we used sparse channel estimation methods, and applied the techniques discussed in Chapters 2 and 3 to design training sequences for these systems. We further exploited some of the structure of this cooperative system, leading to the Support Constrained OMP algorithm to improve the channel estimation and BER performance. Applying the

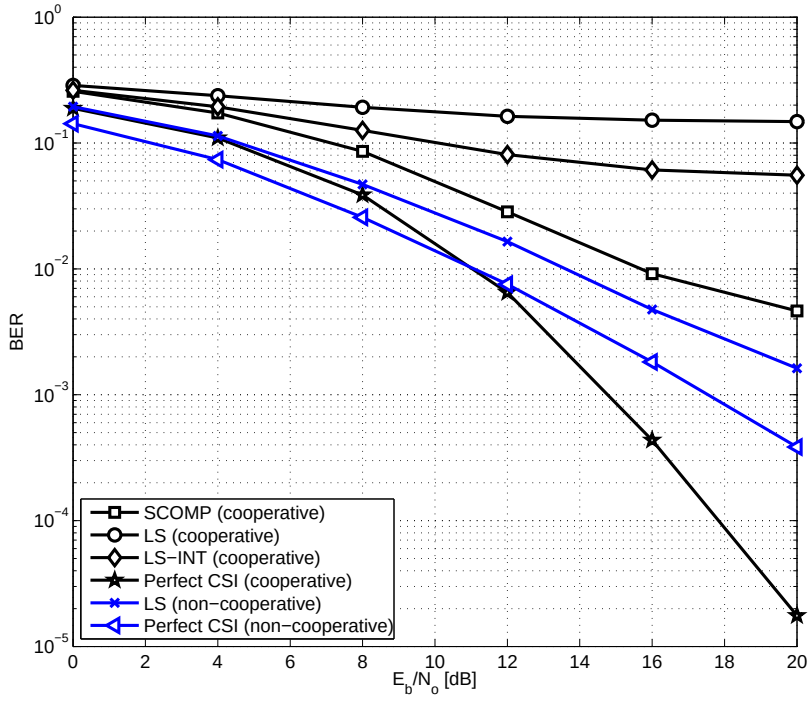

Figure 5.8: Average Probability of Error versus SNR (dB) using TR-STBC $T = 3$ users and $D = 2$ virtual channels. Assumes asynchronous cooperative HF Channel with maximum carrier offset $\xi_{max} = 30$ Hz.

results of Chapters 2 and 3, as well as performing numerical tests, we developed a training design for this cooperative system based on the up-down chirp introduced in Chapter 3. This design offered the best channel estimation performance among the several designs tested.

Through simulation, we demonstrated that the proposed training design, in combination with sparse channel estimation methods, was able to harvest diversity even in the presence of moderate carrier offsets. However as the carrier offsets were increased, the performance gains decreased, and eventually the performance of the scheme was worse than that of the SISO non-cooperative system. As the SISO non-cooperative system had only a single carrier offset between the trans-

mitter and receiver, it was assumed that this offset was correctly compensated for by Carrier tracking methods in the receiver front-end. As such, the only time-variations present in this channel were due to the HF propagation environment, while the channel in the cooperative system consisted of both the HF propagation environment as well as the multiple independent carrier offsets.

CHAPTER 6

CONCLUSIONS

In this thesis we studied the design of training signals for channel estimation in systems that can be modeled as sparse. These designs included the case of asynchronous cooperative communication schemes in OTH-HF propagation environments.

In Chapter 2 based on the conditions for sufficiently informative inputs, we derived training designs for MIMO systems that can be appropriately modeled as sparse. These designs and conditions allowed for improved channel estimation performance in comparison with competing channel estimation schemes that rely on the more restrictive conditions derived from the restricted isometry property and mutual coherence.

In Chapter 3, we further improved upon these training signal designs. This was accomplished by analyzing the performance of an ideal sparse estimation algorithm which led to the introduction of a metric called localized coherence. This metric allowed the comparison of training designs in terms of their channel estimation performance.

In Chapter 4, we considered the asynchronous distributed cooperative communication scheme, first deriving the obtainable diversity and coding gain in frequency-selective channels. This was followed by a comparison of several coding schemes, as well as the effects of the OTH-HF propagation environment. Finally, in Chapter 5 we introduced multiple carrier offsets in the equivalent channel response at the receiver, and proposed a training design for the estimation of these channels. Further exploitation of the structure of the cooperative communication

system led to a modified algorithm for channel estimation. The system was then simulated and it was demonstrated that diversity gains were still achievable for moderate values of carrier offset.

APPENDIX A

APPENDIX: CHAPTER 2

A.1 Proof of Lemma 2.3.1

Assume there are two solutions $\boldsymbol{\alpha}^\star$ and $\boldsymbol{\theta}$ such that $\|\boldsymbol{\alpha}^\star\|_0 = Q$ and $\|\boldsymbol{\theta}\|_0 \leq Q$. Due to the constraint in (2.13), $\boldsymbol{\Upsilon}(\boldsymbol{\alpha}^\star - \boldsymbol{\theta}) = \mathbf{0}$. To prove the lemma, one must show that $(\boldsymbol{\alpha}^\star - \boldsymbol{\theta})$ where $\|\boldsymbol{\alpha}^\star - \boldsymbol{\theta}\|_0 \leq 2S$ is not in the null-space of $\boldsymbol{\Upsilon}$. Since this must hold for all possible $\|\boldsymbol{\alpha}^\star - \boldsymbol{\theta}\| \leq 2S$, we require that every subset of $2S$ columns of $\boldsymbol{\Upsilon}$ is linearly independent, such that the that any vector in the null-space has support greater than $2S$. Note that if the system is not sparse (i.e. $2S \geq R$) we require the matrix $\boldsymbol{\Upsilon}$ to be full column rank.

A.2 Proof of Theorem 2.4.3

Let $\mathbf{D}$ be a $M \times K$ selection matrix, selecting only those samples of $y(c_m T_s)$ corresponding to pilot symbols, $\tilde{\mathbf{s}}$ the vector of these M known pilot symbols, and $\boldsymbol{\Delta}_{\tilde{\mathbf{s}}}$ be the diagonal matrix constructed from $\tilde{\mathbf{s}}$. Defining the diagonal matrix $\{\tilde{\boldsymbol{\Omega}}_{r_f}\}_{m,m} = e^{j2\pi c_m f_{r_f}}$:

$$\tilde{\mathbf{y}} = \boldsymbol{\Delta}_{\tilde{\mathbf{s}}}^{-1}\mathbf{D}\mathbf{y} = \textstyle\sum_{r_f=0}^{R_f-1} \tilde{\boldsymbol{\Omega}}_{r_f}\boldsymbol{\Delta}_{\tilde{\mathbf{s}}}^{-1}\tilde{\mathbf{s}}\alpha_{r_f} = \boldsymbol{\Upsilon}\boldsymbol{\alpha} \tag{A.1}$$

where the projection matrix $\mathbf{P} = \boldsymbol{\Delta}_{\tilde{\mathbf{s}}}^{-1}\mathbf{D}$ and the $M \times R$ matrix $\boldsymbol{\Upsilon}_{m,r_f} = e^{j2\pi c_m f_{r_f}}$ has the form of (2.15). For symbols that are not uniformly spaced, we apply Lemma 2.4.2 using a uniform grid $f_{r_f} = r_f\epsilon_f - d_f$ and require $\epsilon_f R_f \leq 1$ and $1/\epsilon_f$ prime. The shift d_f is the same for all columns of $\boldsymbol{\Upsilon}$ and will not affect identifiability. For

uniformly spaced symbols, $c_m = mN_s$, we use Lemma 2.4.1 and no longer require these restrictions, only that $e^{j2\pi N_s f_{r_f} T_s}$ unique for all r_f and $M \geq 2S$.

A.3 Proof of Theorem 2.4.4

The sampled signal $s(kT_s) = \{\tilde{\mathbf{F}}\tilde{\mathbf{s}}\}_k$ where $\{\tilde{\mathbf{F}}\}_{k,m} = \frac{1}{\sqrt{K}} e^{-j2\pi \frac{c_m k}{K}}$. A delay in the signal $s(kT_s - \tau_{r_\tau}) = \{\tilde{\mathbf{F}}\boldsymbol{\Delta}_{\tilde{\mathbf{s}}}\mathbf{e}_{r_\tau}\}_k$ where $\{\mathbf{e}_{r_\tau}\}_m = e^{j2\pi \frac{c_m \tau_{r_\tau}}{KT_s}}$, and the diagonal matrix $\{\boldsymbol{\Delta}_{\tilde{\mathbf{s}}}\}_{m,m} = \tilde{s}_m$. Taking K observations after removing the cyclic prefix:

$$\mathbf{y} = \sum_{r_\tau=0}^{R_\tau-1} \mathbf{s}_{r_\tau}\alpha_{r_\tau} = \tilde{\mathbf{F}}\boldsymbol{\Delta}_{\tilde{\mathbf{s}}} \sum_{r_\tau=0}^{R_\tau-1} \mathbf{e}_{r_\tau}\alpha_{r_\tau} = \tilde{\mathbf{F}}\boldsymbol{\Delta}_{\tilde{\mathbf{s}}}\boldsymbol{\Upsilon}\boldsymbol{\alpha} \tag{A.2}$$

where $\{\boldsymbol{\Upsilon}\}_{m,r_\tau} = e^{j2\pi \frac{c_m \tau_{r_\tau}}{KT_s}}$. Defining $\mathbf{P} = \boldsymbol{\Delta}_{\tilde{\mathbf{s}}}^{-1}\tilde{\mathbf{F}}^H$, $\tilde{\mathbf{y}} = \mathbf{P}\mathbf{y} = \boldsymbol{\Upsilon}\boldsymbol{\alpha}$. The matrix $\boldsymbol{\Upsilon}$ is similar to (2.15) and thus we require that the elements $e^{j2\pi\tau_{r_\tau}/(KT_s)}$ are distinct. If the M pilots are not uniformly spaced we use Lemma 2.4.2, and assume $\tau_{r_\tau} = r_\tau\epsilon_\tau - d_\tau$, with $\epsilon_\tau R_\tau \leq KT_s$ and KT_s/ϵ_τ prime. If the M pilots are uniformly spaced ($c_m = mN_s$), we use Lemma 2.4.1 and require only the elements $e^{j2\pi N_s\tau_{r_\tau}/(KT_s)}$ to be distinct and $M \geq 2S$.

A.4 Proof of Theorem 2.4.5

The delayed chirp signal $s(kT_s - \tau_{r_\tau}) = e^{-j\pi\frac{\beta}{KT_s^2}(kT_s-\tau_{r_\tau})^2} = \{\mathbf{D}\mathbf{e}_{r_\tau}\{\boldsymbol{\Delta}_\tau\}_{r_\tau,r_\tau}\}_k$. The diagonal matrices $\{\mathbf{D}\}_{k,k} = e^{-j\pi\frac{\beta}{KT_s^2}(kT_s)^2}$ and $\{\boldsymbol{\Delta}_\tau\}_{r_\tau,r_\tau} = e^{-j\pi\frac{\beta}{KT_s^2}(\tau_{r_\tau})^2}$, while the vector $\{\mathbf{e}_{r_\tau}\}_k = e^{j2\pi\frac{\beta}{KT_s}k\tau_{r_\tau}}$. Taking K samples after removing the prefix:

$$\begin{aligned}\mathbf{y} &= \sum_{\mathbf{r}=(r_f,r_\tau)} \boldsymbol{\Omega}_{r_f}\mathbf{s}_{r_\tau}\{\boldsymbol{\alpha}\}_{i(\mathbf{r})} = \mathbf{D}\sum_{\mathbf{r}} \boldsymbol{\Omega}_{r_f}\mathbf{e}_{r_\tau}\{\boldsymbol{\Delta}_\tau\}_{r_\tau,r_\tau}\{\boldsymbol{\alpha}\}_{i(\mathbf{r})} \\ &= \mathbf{D}\textstyle\sum_{\mathbf{r}} \mathbf{e}(f_{r_f} + \frac{\beta}{KT_s}\tau_{r_\tau})\{\boldsymbol{\Delta}_\tau\}_{r_\tau,r_\tau}\{\boldsymbol{\alpha}\}_{i(\mathbf{r})} = \mathbf{D}\boldsymbol{\Upsilon}(\boldsymbol{\Delta}_\tau \otimes \mathbf{I})\boldsymbol{\alpha}\end{aligned} \tag{A.3}$$

where $\{\boldsymbol{\Upsilon}\}_{k,r_\tau+R_\tau r_f} = e^{j2\pi k(f_{r_f}+\frac{\beta}{KT_s}\tau_{r_\tau})}$. Note $(\mathbf{I}\otimes\boldsymbol{\Delta}_\tau)\boldsymbol{\alpha}$ has the same sparsity as $\boldsymbol{\alpha}$ since $(\mathbf{I}\otimes\boldsymbol{\Delta}_\tau)$ is a full-rank diagonal matrix. Defining the matrix $\mathbf{P}=\mathbf{D}^H$, the proof follows using Lemma 2.4.1.

A.5 Proof of Theorem 2.4.6

By design, the samples of the vth transmit waveform $s_v(kT_s)=\delta[v-k]\tilde{s}_v$, implying $\mathbf{S}=\boldsymbol{\Delta}_{\tilde{\mathbf{s}}}$.

$$\mathbf{y}=\sum_{\mathbf{r}=(r_\mu,r_f)}\boldsymbol{\Omega}_{r_f}\boldsymbol{\Delta}_{\tilde{\mathbf{s}}}\mathbf{e}_{r_\mu}\{\boldsymbol{\alpha}\}_{i(\mathbf{r})}=\boldsymbol{\Delta}_{\tilde{\mathbf{s}}}\boldsymbol{\Upsilon}\boldsymbol{\alpha} \tag{A.4}$$

where $\{\boldsymbol{\Upsilon}\}_{k,r_f+r_\mu R_f} = e^{j2\pi k(f_{r_f}+\mu_{r_\mu})}$. Supposing the elements $e^{j2\pi(f_{r_f}+\mu_{r_\mu})}$ are unique, and defining the projection matrix $\mathbf{P}=\boldsymbol{\Delta}_{\tilde{\mathbf{s}}}^{-1}$, the proof follows via Lemma 2.4.1 with the condition $N_{tx}=K\geq 2S$.

A.6 Proof of Theorem 2.4.7

By the design of $s_v(t)$ the $K\times N_{tx}$ matrix $\mathbf{S}_{r_\tau}=\mathbf{F}\boldsymbol{\Omega}_{r_\tau}$ where the $K\times N_{tx}$ matrix $\{\mathbf{F}\}_{k,v}=e^{-j2\pi\frac{kv}{K}}$ and the $N_{tx}\times N_{tx}$ diagonal matrix $\{\boldsymbol{\Omega}_{r_\tau}\}_{v,v}=e^{j2\pi\frac{v\tau_{r_\tau}}{KT_s}}$.

$$\mathbf{y}=\sum_{\mathbf{r}=(r_\mu,r_\tau)}\mathbf{S}_{r_\tau}\mathbf{e}_{r_\mu}\{\boldsymbol{\alpha}\}_{i(\mathbf{r})}=\mathbf{F}\sum_{\mathbf{r}}\boldsymbol{\Omega}_{r_\tau}\mathbf{e}_{r_\mu}\{\boldsymbol{\alpha}\}_{i(\mathbf{r})}=\mathbf{F}\boldsymbol{\Upsilon}\boldsymbol{\alpha} \tag{A.5}$$

Defining the projection matrix $\mathbf{P}=K^{-1}\mathbf{F}^H$, $\mathbf{Py}=\boldsymbol{\Upsilon}\boldsymbol{\alpha}$ with $\{\boldsymbol{\Upsilon}\}_{v,r_\tau+r_\mu R_\tau} = e^{j2\pi v(\mu_{r_\mu}+\frac{\tau_{r_\tau}}{KT_s})}$. Supposing the elements $e^{j2\pi(\mu_{r_\mu}+\frac{\tau_{r_\tau}}{KT_s})}$ are unique, the proof follows via Lemma 2.4.1 with the condition $N_{tx}\geq 2S$.

A.7 Proof of Theorem 2.4.8

The $N_{rx} \times R_\nu$ matrix $\{\mathbf{M}_{out}\}_{u,r_\nu} = e^{j2\pi u \nu_{r_\nu}}$ has the structure in (2.15). As we assume a uniform linear array, we make use of Lemma 2.4.1 and require that all the elements $e^{j2\pi\nu_{r_\nu}}$ are unique and $N_{rx} \geq 2S$. Note that if the array elements are not uniformly spaced, assuming a uniform quantization $\nu_{r_\nu} = r_\nu \epsilon_\nu - d_\nu$ one can use Lemma 2.4.2, requiring a prime $1/\epsilon_\nu$.

A.8 Proof of Theorem 2.4.9

Assuming S arrival angles, the observations $\mathbf{Y}^T = \mathbf{M}_{out}(\mathcal{S})\boldsymbol{\mathcal{A}}^\star_{\mathcal{S}}\mathbf{M}_{in}^T$ where $\boldsymbol{\mathcal{A}}^\star_{\mathcal{S}}$ consists of the subset of rows from $\boldsymbol{\mathcal{A}}^\star$ with nonzero elements and $\mathbf{M}_{out}(\mathcal{S})$ are the columns of $\mathbf{M}_{out}$ corresponding to the support $\mathcal{S}$ previously found using (2.23). Define $\tilde{\mathbf{Y}} = \tilde{\mathbf{P}}(\mathbf{M}^\dagger_{out}(\mathcal{S})\mathbf{Y}^T)^T = \tilde{\mathbf{P}}\mathbf{M}_{in}(\boldsymbol{\mathcal{A}}^\star_{\mathcal{S}})^T$. Since by assumption each angle of arrival corresponds to a particular propagation path, each of the columns of $(\boldsymbol{\mathcal{A}}^\star_{\mathcal{S}})^T$ contains only one nonzero element corresponding to the particular angle of departure, doppler and delay of this path (μ, f, τ). Thus the s-th column $[\tilde{\mathbf{Y}}]_s = \tilde{\mathbf{P}}\mathbf{M}_{in}\boldsymbol{\alpha}^\star_s$ where $\boldsymbol{\alpha}^\star_s$ is the s-th column of $(\boldsymbol{\mathcal{A}}^\star_{\mathcal{S}})^T$ which has only one non-zero element and via Lemma 2.3.1 (replacing $\boldsymbol{\Upsilon}$ with $\tilde{\mathbf{P}}\mathbf{M}_{in}$ and $\boldsymbol{\alpha}^\star$ with $\boldsymbol{\alpha}^\star_s$) the solution is unique if every pair of columns of $\tilde{\mathbf{P}}\mathbf{M}_{in}$ is linearly independent.

APPENDIX B

APPENDIX: CHAPTER 4

B.1 Derivation of CRLB

We begin with the derivation of the CRLB for a generic system model consisting of the superposition of a fixed number of signals, followed by the application to the specific signal model of interest.

B.1.1 CRLB for Superposition of Signals

The system model consists of a finite known number Q of superimposed signals, each a function of a set of parameters $\boldsymbol{\kappa}_q$:

$$\mathbf{y} = \sum_{q=0}^{Q-1} (\boldsymbol{\kappa}_q) a_q + \mathbf{w} = \mathbf{M}(\boldsymbol{\kappa})\mathbf{a} \tag{B.1}$$

where $\mathbf{w} \sim \mathcal{CN}(0, \mathbf{I}\sigma_w^2)$, $\mathbf{y} = [y[0], ..., y[K-1]]^T$, $\mathbf{M}(\boldsymbol{\kappa}) \in \mathbb{C}^{K \times Q}$, and $\mathbf{a} \in \mathbb{C}^Q$ where in general $\boldsymbol{\kappa}_q = (\kappa_{1,q}, ..., \kappa_{N,q})^T$. Further, we define the vector $\boldsymbol{\eta}_n = (\kappa_{n,1}, ..., \kappa_{n,Q})^T$. We are interested in computing the CRLB with respect to the parameters $\boldsymbol{\xi} = \{\mathbf{a}, \mathbf{a}^*, \boldsymbol{\eta}_1, ..., \boldsymbol{\eta}_N\}$, where we assume the parameters $\boldsymbol{\kappa}$ (and thus β) are real parameters. The Fisher Information matrix $\mathbf{I}(\boldsymbol{\kappa})$ is given by:

$$\mathbf{I}(\boldsymbol{\xi}) = \mathbb{E}\left\{ \frac{\partial}{\partial \boldsymbol{\xi}^*} \ln p(\mathbf{y}|\boldsymbol{\xi}) \left(\frac{\partial}{\partial \boldsymbol{\xi}^*} \ln p(\mathbf{y}|\boldsymbol{\xi}) \right)^H \right\} \tag{B.2}$$

and for any unbiased estimate of the parameters $\boldsymbol{\xi}$, $\mathrm{Cov}(\boldsymbol{\xi}) \geq \mathbf{I}^{-1}(\boldsymbol{\xi})$. Note that the observations have a complex Gaussian PDF or:

$$p(\mathbf{y}|\boldsymbol{\xi}) = \frac{1}{\pi^K (\sigma_w^2)^K} \exp\left(-\frac{\|\mathbf{y} - \mathbf{M}(\boldsymbol{\kappa})\mathbf{a}\|^2}{\sigma_w^2} \right) \tag{B.3}$$

In the following, to simplify notation we denote $\mathbf{M} \triangleq \mathbf{M}(\boldsymbol{\kappa})$. Under these assumptions, the gradients:

$$\frac{\partial}{\partial \mathbf{a}^*}\|\mathbf{y}-\mathbf{M}\mathbf{a}\|^2 = -\mathbf{M}^H(\mathbf{y}-\mathbf{M}\mathbf{a}) \tag{B.4}$$

$$\frac{\partial}{\partial \mathbf{a}}\|\mathbf{y}-\mathbf{M}\mathbf{a}\|^2 = -\mathbf{M}^T(\mathbf{y}-\mathbf{M}\mathbf{a})^* \tag{B.5}$$

$$\frac{\partial}{\partial \boldsymbol{\eta}_n}\|\mathbf{y}-\mathbf{M}\mathbf{a}\|^2 = -2\Re\{\tilde{\mathbf{A}}^H\mathbf{D}_{\boldsymbol{\eta}_n}^H(\mathbf{y}-\mathbf{M}\mathbf{a})\} \tag{B.6}$$

where $\tilde{\boldsymbol{\mathcal{A}}} = \mathrm{diag}(\mathbf{a})$ and $\{\mathbf{D}_{\boldsymbol{\eta}_n}\}_{k,q} = \frac{\partial\{(\boldsymbol{\kappa}_q)\}_k}{\partial \boldsymbol{\kappa}_{n,q}}$, where $(\boldsymbol{\kappa}_q)$ corresponds to the q-th column of $\mathbf{M}$. Stacking the partial derivatives w.r.t. $\boldsymbol{\kappa} = (\boldsymbol{\eta}_1^T, ..., \boldsymbol{\eta}_N^T)^T$:

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\kappa}}\|\mathbf{y}-\mathbf{M}\mathbf{a}\|^2 &= \begin{bmatrix} \frac{\partial}{\partial \boldsymbol{\eta}_1}\|\mathbf{y}-\mathbf{M}\mathbf{a}\|^2 \\ \vdots \\ \frac{\partial}{\partial \boldsymbol{\eta}_N}\|\mathbf{y}-\mathbf{M}\mathbf{a}\|^2 \end{bmatrix} \\
&= \begin{bmatrix} -2\Re\left(\tilde{\boldsymbol{\mathcal{A}}}^H\mathbf{D}_{\boldsymbol{\eta}_1}^H(\mathbf{y}-\mathbf{M}\mathbf{a})\right) \\ \vdots \\ -2\Re\left(\tilde{\boldsymbol{\mathcal{A}}}^H\mathbf{D}_{\boldsymbol{\eta}_N}^H(\mathbf{y}-\mathbf{M}\mathbf{a})\right) \end{bmatrix} \qquad \text{(B.7)} \\
&= -2\Re\{\boldsymbol{\mathcal{A}}^H\mathbf{D}^H(\mathbf{y}-\mathbf{M}\boldsymbol{\mathcal{A}})\} \qquad \text{(B.8)}
\end{aligned}$$

where $\mathbf{D} \triangleq [\mathbf{D}_{\boldsymbol{\eta}_1}, ..., \mathbf{D}_{\boldsymbol{\eta}_N}]$ and $\boldsymbol{\mathcal{A}} \triangleq \mathbf{I}_N \otimes \tilde{\boldsymbol{\mathcal{A}}}$. With a simple calculation, the Fisher information matrix can be expressed as:

$$\mathbf{I}(\boldsymbol{\kappa}) = \frac{1}{\sigma_w^2}\begin{bmatrix} \mathbf{M}^H\mathbf{M} & \mathbf{0} & \mathbf{M}^H\mathbf{D}\boldsymbol{\mathcal{A}} \\ \mathbf{0} & \mathbf{M}^T\mathbf{M}^* & \mathbf{M}^T\mathbf{D}^*\boldsymbol{\mathcal{A}}^* \\ \boldsymbol{\mathcal{A}}^H\mathbf{D}^H\mathbf{M} & \boldsymbol{\mathcal{A}}^T\mathbf{D}^T\mathbf{M}^* & 2\Re\{\boldsymbol{\mathcal{A}}^H\mathbf{D}^H\mathbf{D}\boldsymbol{\mathcal{A}}\} \end{bmatrix} \tag{B.9}$$

To compute the inverse of $\mathbf{I}(\boldsymbol{\kappa})$, recall the inverse of a partitioned Hermitian matrix:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^H & \mathbf{C} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{A}-\mathbf{B}\mathbf{C}^{-1}\mathbf{B}^H)^{-1} & \mathbf{A}^{-1}\mathbf{B}(\mathbf{B}^H\mathbf{A}^{-1}\mathbf{B}-\mathbf{C})^{-1} \\ (\mathbf{B}^H\mathbf{A}^{-1}\mathbf{B}-\mathbf{C})^{-1}\mathbf{B}^H\mathbf{A}^{-1} & (\mathbf{C}-\mathbf{B}^H\mathbf{A}^{-1}\mathbf{B})^{-1} \end{bmatrix} \tag{B.10}$$

where we assume the appropriate matrices are nonsingular. In the following, we map the portions of $\mathbf{I}(\boldsymbol{\kappa})$ to to a block diagonal Hermitian matrix as follows:

$$\begin{aligned}
\mathbf{A} &= \begin{bmatrix} \mathbf{M}^H\mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}^T\mathbf{M}^* \end{bmatrix} \\
\mathbf{B} &= \begin{bmatrix} \mathbf{M}^H\mathbf{D}\boldsymbol{\mathcal{A}} \\ \mathbf{M}^T\mathbf{D}^*\boldsymbol{\mathcal{A}}^* \end{bmatrix} \\
\mathbf{C} &= 2\Re\{\boldsymbol{\mathcal{A}}^H\mathbf{D}^H\mathbf{D}\boldsymbol{\mathcal{A}}\}
\end{aligned} \tag{B.11}$$

In the following we will analyze the individual blocks constituting the inverse of the Fisher information matrix and corresponding to the appropriate components of the CRB

CRLB for $\boldsymbol{\kappa}$

We will use the notation $\mathrm{CRB}(\boldsymbol{\kappa})$ to denote the corresponding block of the inverse of the Fisher information matrix, $\mathrm{Cov}(\boldsymbol{\kappa}) \geq \mathrm{CRB}(\boldsymbol{\kappa})$. We are interested in the bottom right hand corner of the inverse of $\mathbf{I}(\boldsymbol{\kappa})$, and using the inverse of a block Hermitian defined above $\mathrm{CRB}(\boldsymbol{\kappa}) = (\mathbf{C} - \mathbf{B}^H\mathbf{A}^{-1}\mathbf{B})^{-1}$. Substitution of the appropriate matrices gives:

$$\begin{aligned}
\mathrm{CRB}(\boldsymbol{\kappa}) &= \sigma_w^2 \Bigg(2\Re\{\boldsymbol{\mathcal{A}}^H\mathbf{D}^H\mathbf{D}\boldsymbol{\mathcal{A}}\} - \begin{bmatrix} \boldsymbol{\mathcal{A}}^H\mathbf{D}^H\mathbf{M} & \boldsymbol{\mathcal{A}}^T\mathbf{D}^T\mathbf{M}^* \end{bmatrix} \\
&\quad \times \begin{bmatrix} (\mathbf{M}^H\mathbf{M})^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{M}^T\mathbf{M}^*)^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{M}^H\mathbf{D}\boldsymbol{\mathcal{A}} \\ \mathbf{M}^T\mathbf{D}^*\boldsymbol{\mathcal{A}}^* \end{bmatrix} \Bigg)^{-1} && \text{(B.12)} \\
&= \sigma_w^2 \left(2\Re\{\boldsymbol{\mathcal{A}}^H\mathbf{D}^H\mathbf{D}\boldsymbol{\mathcal{A}}\} - 2\Re\{\boldsymbol{\mathcal{A}}^H\mathbf{D}^H\mathbf{M}(\mathbf{M}^H\mathbf{M})^{-1}\mathbf{M}^H\mathbf{D}\boldsymbol{\mathcal{A}}\} \right)^{-1} && \text{(B.13)} \\
&= \sigma_w^2 \left(2\Re\{\boldsymbol{\mathcal{A}}^H\mathbf{D}^H(\mathbf{I} - \mathbf{M}(\mathbf{M}^H\mathbf{M})^{-1}\mathbf{M}^H)\mathbf{D}\boldsymbol{\mathcal{A}}\} \right)^{-1} && \text{(B.14)}
\end{aligned}$$

CRLB for a

To compute CRB($\mathbf{a}$) we first take the top left hand portion of $\mathbf{I}^{-1}(\boldsymbol{\kappa})$ such that CRB$(\mathbf{a}, \mathbf{a}^*) = (\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^H)^{-1}$. After the appropriate substitution:

$$\begin{aligned} \text{CRB}(\mathbf{a}, \mathbf{a}^*) &= \sigma_w^2 \Bigg(\begin{bmatrix} \mathbf{M}^H\mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}^T\mathbf{M}^* \end{bmatrix} - \begin{bmatrix} \mathbf{M}^H\mathbf{D}\boldsymbol{\mathcal{A}} \\ \mathbf{M}^T\mathbf{D}^*\boldsymbol{\mathcal{A}}^* \end{bmatrix} \\ &\quad \times (2\Re\{\boldsymbol{\mathcal{A}}^H\mathbf{D}^H\mathbf{D}\boldsymbol{\mathcal{A}}\})^{-1} \begin{bmatrix} \boldsymbol{\mathcal{A}}^H\mathbf{D}^H\mathbf{M} & \boldsymbol{\mathcal{A}}^T\mathbf{D}^T\mathbf{M}^* \end{bmatrix} \Bigg)^{-1} \end{aligned} \tag{B.15}$$

We now make use of the Woodbury matrix identity:

$$(\mathbf{W} + \mathbf{V}\mathbf{U}\mathbf{Z})^{-1} = \mathbf{W}^{-1} - \mathbf{W}^{-1}\mathbf{V}(\mathbf{U}^{-1} + \mathbf{Z}\mathbf{W}^{-1}\mathbf{V})^{-1}\mathbf{Z}\mathbf{W}^{-1} \tag{B.16}$$

$$= \mathbf{W}^{-1}(\mathbf{I} - \mathbf{V}(\mathbf{U}^{-1} + \mathbf{Z}\mathbf{W}^{-1}\mathbf{V})^{-1}\mathbf{Z}\mathbf{W}^{-1}) \tag{B.17}$$

where with respect to our system the matrices are defined as follows:

$$\begin{aligned} \mathbf{W} &= \begin{bmatrix} \mathbf{M}^H\mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}^T\mathbf{M}^* \end{bmatrix} \\ \mathbf{V} &= -\begin{bmatrix} \mathbf{M}^H\mathbf{D}\boldsymbol{\mathcal{A}} \\ \mathbf{M}^T\mathbf{D}^*\boldsymbol{\mathcal{A}}^* \end{bmatrix} \\ \mathbf{U} &= (2\Re\{\boldsymbol{\mathcal{A}}^H\mathbf{D}^H\mathbf{D}\boldsymbol{\mathcal{A}}\})^{-1} \\ \mathbf{Z} &= \begin{bmatrix} \boldsymbol{\mathcal{A}}^H\mathbf{D}^H\mathbf{M} & \boldsymbol{\mathcal{A}}^T\mathbf{D}^T\mathbf{M}^* \end{bmatrix} \end{aligned} \tag{B.18}$$

We can now express CRB$(\mathbf{a}, \mathbf{a}^*)$ as:

$$\begin{aligned} \text{CRB}(\mathbf{a}, \mathbf{a}^*) &= \sigma_w^2 \Bigg(\begin{bmatrix} (\mathbf{M}^H\mathbf{M})^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{M}^T\mathbf{M}^*)^{-1} \end{bmatrix} + \begin{bmatrix} (\mathbf{M}^H\mathbf{M})^{-1}\mathbf{M}^H\mathbf{D}\boldsymbol{\mathcal{A}} \\ (\mathbf{M}^T\mathbf{M}^*)^{-1}\mathbf{M}^T\mathbf{D}^*\boldsymbol{\mathcal{A}}^* \end{bmatrix} \\ &\quad \times \frac{\text{CRB}(\kappa)}{\sigma_w^2} \begin{bmatrix} \boldsymbol{\mathcal{A}}^H\mathbf{D}^H\mathbf{M}(\mathbf{M}^H\mathbf{M})^{-1} & \boldsymbol{\mathcal{A}}^T\mathbf{D}^T\mathbf{M}^*(\mathbf{M}^T\mathbf{M}^*)^{-1} \end{bmatrix} \Bigg)^{-1} \end{aligned} \tag{B.19}$$

As we are only interested in the top left hand corner of the matrix $\mathrm{CRB}(\mathbf{a}, \mathbf{a}^*)$, we have:

$$\mathrm{Cov}(\mathbf{a}) \geq \mathrm{CRB}(\mathbf{a}) = \sigma_w^2(\mathbf{M}^H\mathbf{M})^{-1} + (\mathbf{M}^H\mathbf{M})^{-1}\mathbf{M}^H\mathbf{D}\boldsymbol{\mathcal{A}}\mathrm{CRB}(\boldsymbol{\kappa})\boldsymbol{\mathcal{A}}^H\mathbf{D}^H\mathbf{M}(\mathbf{M}^H\mathbf{M})^{-1} \tag{B.20}$$

B.1.2 CRLB for Transformations of Parameters

Suppose that we could decompose the matrix $\mathbf{M}(\boldsymbol{\kappa}) = \mathbf{X}\mathbf{G}(\boldsymbol{\kappa})$ where $\mathbf{X}$ is a known matrix, and we are interested in estimating $\mathbf{h} = \mathbf{G}(\boldsymbol{\kappa})\mathbf{a}$. We can obtain the CRB for the parameter $\mathbf{h}$ as:

$$\mathrm{Cov}(\hat{\mathbf{h}}) \geq \mathbf{J}\mathbf{I}^{-1}(\boldsymbol{\xi})\mathbf{J} \tag{B.21}$$

where $\{\mathbf{J}\}_{m,n} = \frac{\partial\{\mathbf{G}(\boldsymbol{\kappa})\mathbf{a}\}_m}{\partial\xi_n}$ is the Jacobian of the transformation $\mathbf{G}(\boldsymbol{\kappa})\mathbf{a}$ with respect to the parameters $\boldsymbol{\xi} = (\mathbf{a}, \mathbf{a}^*, \boldsymbol{\kappa})$.

Proof. Similar to the proof of the CRLB, define the transformation of the parameters $\mathbf{g}(\boldsymbol{\xi})$ and its estimate as $\hat{\mathbf{g}}(\boldsymbol{\xi})$, define $\mathbf{s}(\mathbf{y};\boldsymbol{\xi}) = \bigtriangledown_{\boldsymbol{\xi}^*} \ln p(\mathbf{y};\boldsymbol{\xi})$ as the scoring function and $\mathbf{z} = [\mathbf{s}(\mathbf{y};\boldsymbol{\xi})^T, (\hat{\mathbf{g}}(\boldsymbol{\xi}) - \mathbf{g}(\boldsymbol{\xi}))^T]^T$. Then:

$$\mathbb{E}\{\mathbf{z}\mathbf{z}^H\} = \begin{bmatrix} \mathbf{I}(\boldsymbol{\xi}) & \mathbb{E}(\mathbf{s}(\mathbf{y};\boldsymbol{\xi})(\hat{\mathbf{g}}(\boldsymbol{\xi}) - \mathbf{g}(\boldsymbol{\xi}))^H) \\ \mathbb{E}((\hat{\mathbf{g}}(\boldsymbol{\xi}) - \mathbf{g}(\boldsymbol{\xi}))\mathbf{s}^H(\mathbf{y};\boldsymbol{\xi})) & \mathrm{Cov}(\hat{\mathbf{g}}(\boldsymbol{\xi})) \end{bmatrix} \tag{B.22}$$

Note that due to the regularity conditions:

$$\begin{aligned} \mathbb{E}\{\mathbf{s}(\mathbf{y};\boldsymbol{\xi})(\hat{\mathbf{g}}(\boldsymbol{\xi}) - \mathbf{g}(\boldsymbol{\xi}))^H\}_{i,j} &= \int (\hat{\mathbf{g}}_j^*(\boldsymbol{\xi}) - \mathbf{g}_j^*(\boldsymbol{\xi}))p(\mathbf{y};\boldsymbol{\xi})\frac{\partial}{\partial\xi_i^*}p(\mathbf{y};\boldsymbol{\xi})d\mathbf{y} \\ &= \int \hat{\mathbf{g}}_j^*(\boldsymbol{\xi})\frac{\partial}{\partial\xi_i^*}p(\mathbf{y};\boldsymbol{\xi})d\mathbf{y} \\ &= \frac{\partial\mathbf{g}_j^*(\boldsymbol{\xi})}{\partial\xi_i^*} \end{aligned} \tag{B.23}$$

Likewise, it can be shown $\mathbb{E}\{(\hat{\mathbf{g}}(\boldsymbol{\xi}) - \mathbf{g}(\boldsymbol{\xi}))\mathbf{s}^H(\mathbf{y};\boldsymbol{\xi})\}_{i,j} = \frac{\partial \mathbf{g}_i(\boldsymbol{\xi})}{\partial \xi_j} = \{\mathbf{J}\}_{i,j}$, and we define the Jacobian $\mathbf{J}$ as:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial}{\partial \xi_1}\mathbf{g}_1(\boldsymbol{\xi}) & \cdots & \frac{\partial}{\partial \xi_N}\mathbf{g}_1(\boldsymbol{\xi}) \\ \vdots & \cdots & \vdots \\ \frac{\partial}{\partial \xi_1}\mathbf{g}_M(\boldsymbol{\xi}) & \cdots & \frac{\partial}{\partial \xi_N}\mathbf{g}_M(\boldsymbol{\xi}) \end{bmatrix} \tag{B.24}$$

Making use of the block diagonalization of a matrix and the Decorrelation property, one can show:

$$\mathrm{Cov}(\hat{\mathbf{g}}(\boldsymbol{\xi})) - \mathbf{J}\mathbf{I}^{-1}(\boldsymbol{\xi})\mathbf{J}^H \geq \mathbf{0} \tag{B.25}$$

□

B.1.3 CRLB for MIMO Model

The model of interest is simply a superposition of Q signals and can be expressed in the form Note that this model is simply a superposition of Q signals and can thus be expressed in the form (B.1).

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{w} = \mathbf{X}\mathbf{G}(\boldsymbol{\kappa})\mathbf{a} + \mathbf{w} = \mathbf{M}(\boldsymbol{\kappa})\mathbf{a} + \mathbf{w} \tag{B.26}$$

where we recall

$$\{\mathbf{G}(\boldsymbol{\kappa})\}_{l+kL+vLK+uLKN_{tx},q} = e^{j2\pi(\omega_q k+\nu_q u+\mu_q v)}p(lT_s - \tau_q) \tag{B.27}$$

$$\{\mathbf{M}(\boldsymbol{\kappa})\}_{k+uK,q} = \sum_{v=0}^{N_{tx}-1}\sum_{l=0}^{L-1} x_v[k-l]e^{j2\pi(\omega_q k+\nu_q u+\mu_q v)}p(lT_s - \tau_q) \tag{B.28}$$

$$= \sum_{v=0}^{N_{tx}-1} e^{j2\pi(\omega_q k+\nu_q u+\mu_q v)}s_v(kT_s - \tau_q) \tag{B.29}$$

With respect to the parameter set[1] $\boldsymbol{\xi} = \{\mathbf{a}, \mathbf{a}^*, \boldsymbol{\nu}, \boldsymbol{\mu}, \boldsymbol{\omega}, \boldsymbol{\tau}\}$, we first must calculate the matrix $\mathbf{D} = [\mathbf{D}_{\boldsymbol{\nu}}, \mathbf{D}_{\boldsymbol{\mu}}, \mathbf{D}_{\boldsymbol{\omega}}, \mathbf{D}_{\boldsymbol{\tau}}]$. Recall $\{\mathbf{D}_{\boldsymbol{\nu}}\}_{k,q} = \frac{\partial\{(\boldsymbol{\kappa}_q)\}_k}{\partial \nu_q}$, where $(\boldsymbol{\kappa}_q)$ is the

[1]Note that we have made the following substitution $(\boldsymbol{\eta}_1, ..., \boldsymbol{\eta}_N) = (\boldsymbol{\nu}, \boldsymbol{\mu}, \boldsymbol{\omega}, \boldsymbol{\tau})$. Further when calculating the matrices $\mathbf{D}_{\boldsymbol{\eta}_n}$, we replace $\boldsymbol{\eta}_n$ with the appropriate parameter vector.

q-th column of the matrix $\mathbf{M}(\boldsymbol{\kappa})$, we thus have:

$$\begin{aligned}
\{\mathbf{D}_{\boldsymbol{\nu}}\}_{k+uK,q} &= j2\pi u\{\mathbf{M}\}_{k+uK,q} \\
\{\mathbf{D}_{\boldsymbol{\mu}}\}_{k+uK,q} &= e^{j2\pi(\omega_q k+\nu_q u)} \sum_{v=0}^{N_{tx}-1} (j2\pi v) e^{j2\pi\mu_q v} s_v(kT_s-\tau_q) \\
\{\mathbf{D}_{\boldsymbol{\omega}}\}_{k+uK,q} &= j2\pi k\{\mathbf{M}\}_{k+uK,q} \\
\{\mathbf{D}_{\boldsymbol{\tau}}\}_{k+uK,q} &= \frac{\partial\{\mathbf{M}\}_{k+uK,p}}{\partial\tau_q} = e^{j2\pi(\omega_q k+\nu_q u)} \sum_{v=0}^{N_{tx}-1} e^{j2\pi\mu_q v} \frac{\partial s_v(kT_s-\tau_q)}{\partial\tau_q}
\end{aligned} \tag{B.30}$$

To get the CRB for the parameters $\mathbf{a}$ and $\boldsymbol{\kappa}$, we now have defined all the elements necessary to substitute into the expressions (B.12) and (B.20). However, our ultimate goal is to compute the CRB for the transformation of the estimates $\hat{\boldsymbol{\kappa}}$ into an estimate of the channel impulse response $\mathbf{h}$. To achieve this we need to compute the Jacobian $\mathbf{J}$ of transformation $\mathbf{h} = \mathbf{G}(\boldsymbol{\kappa})\mathbf{a}$ as discussed previously. It is easily shown that the matrix $\mathbf{J} = [\mathbf{G}, \mathbf{0}, \mathbf{G}'\mathbf{A}]$, where the matrix $\mathbf{G}' = [\mathbf{G}'_{\boldsymbol{\nu}}, \mathbf{G}'_{\boldsymbol{\mu}}, \mathbf{G}'_{\boldsymbol{\omega}}, \mathbf{G}'_{\boldsymbol{\tau}}]$. Similar to the matrix $\mathbf{D}_{\boldsymbol{\nu}}$, the matrices $\{\mathbf{G}'_{\boldsymbol{\nu}}\}_{z,q} = \frac{\partial\{\mathbf{g}(\boldsymbol{\kappa}_q)\}_k}{\partial\nu_q}$ where $\mathbf{g}(\boldsymbol{\kappa}_q)$ is the q-th column of $\mathbf{G}(\boldsymbol{\kappa})$. The particular matrices are:

$$\begin{aligned}
\{\mathbf{G}'_{\boldsymbol{\nu}}\}_{l+kL+vLK+uLKN_{tx},q} &= j2\pi u\{\mathbf{G}\}_{l+kL+vLK+uLKN_{tx},q} \\
\{\mathbf{G}'_{\boldsymbol{\mu}}\}_{l+kL+vLK+uLKN_{tx},q} &= j2\pi v\{\mathbf{G}\}_{l+kL+vLK+uLKN_{tx},q} \\
\{\mathbf{G}'_{\boldsymbol{\omega}}\}_{l+kL+vLK+uLKN_{tx},q} &= j2\pi k\{\mathbf{G}\}_{l+kL+vLK+uLKN_{tx},q} \\
\{\mathbf{G}'_{\boldsymbol{\tau}}\}_{l+kL+vLK+uLKN_{tx},q} &= e^{j2\pi(\omega_q k+\nu_q u+\mu_q v)} \frac{\partial p(lT_s-\tau_q)}{\partial\tau_q}
\end{aligned}$$

The final step is to compute (B.21). In order to calculate this, we need the lower left and top right corner of the matrix $\mathbf{I}^{-1}(\boldsymbol{\kappa})$. Using (B.10), and after some

algebra:

$$
\begin{aligned}
Cov(\mathbf{h}) &\geq \mathbf{J}\mathbf{I}^{-1}(\boldsymbol{\kappa})\mathbf{J}^H && \text{(B.31)}\\
&= \begin{bmatrix} \mathbf{G} & \mathbf{0} & \mathbf{G}'\boldsymbol{\mathcal{A}} \end{bmatrix} \mathbf{I}^{-1}(\boldsymbol{\kappa}) \begin{bmatrix} \mathbf{G} \\ \mathbf{0} \\ \mathbf{G}'\boldsymbol{\mathcal{A}} \end{bmatrix} && \text{(B.32)}\\
&= \mathbf{G}\mathrm{CRB}(\mathbf{a})\mathbf{G}^H + \mathbf{G}'\boldsymbol{\mathcal{A}}\mathrm{CRB}(\boldsymbol{\kappa})(\mathbf{G}'\boldsymbol{\mathcal{A}})^H \\
&\quad -(\mathbf{G}'\boldsymbol{\mathcal{A}})\mathrm{CRB}(\boldsymbol{\kappa})\boldsymbol{\mathcal{A}}^H\mathbf{D}^H\mathbf{M}(\mathbf{M}^H\mathbf{M})^{-1}\mathbf{G}^H \\
&\quad +\mathbf{G}(\mathbf{M}^H\mathbf{M})^{-1}\mathbf{M}^H\mathbf{D}\boldsymbol{\mathcal{A}}CRB(\boldsymbol{\kappa})(\mathbf{G}'\boldsymbol{\mathcal{A}})^H && \text{(B.33)}
\end{aligned}
$$

APPENDIX C

APPENDIX: CHAPTER 5

C.1 Proof of Theorem 4.3.1

In the following let $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ be the eigenvalue decomposition of $\mathbf{\Phi}_h^{H/2}\boldsymbol{\mathcal{R}}^H\boldsymbol{\mathcal{R}}\mathbf{\Phi}_h^{1/2}$, where $\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_\eta)$ are the nonzero ordered eigenvalues. Further, $\mathbf{V}\mathbf{\Gamma}\mathbf{V}^H$ is the eigenvalue decomposition of $\mathbf{\Delta}_h$ with $\mathbf{\Gamma} = \text{diag}(\gamma_1, \ldots, \gamma_{\eta_L})$ are the ordered nonzero eigenvalues. It can shown that $\mathbf{\Psi} = (\mathbf{\Lambda} \otimes \mathbf{\Gamma})$ corresponds to the nonzero eigenvalues of $\mathbf{Z}^H\mathbf{Z}$. The lower bound is obtained from the lower bound provided in the proof of Lemma 1 in [82]:

$$P_e(SNR|\boldsymbol{\mathcal{R}}) \geq \frac{1}{|M|} Q(\max_{ij}(\sqrt{\alpha_1^{i,j}/2}))P_{deep}(SNR|\boldsymbol{\mathcal{R}}) \tag{C.1}$$

where the probability of the deep fade event:

$$P_{deep}(SNR|\boldsymbol{\mathcal{R}}) = \Pr(\|\mathbf{Z}\tilde{\mathbf{h}}\|^2 \leq 1/SNR|\boldsymbol{\mathcal{R}}) \tag{C.2}$$

Recall the definition of $\mathbf{Z}$ and using the property $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC} \otimes \mathbf{BD})$ we have:

$$\begin{aligned} \|\mathbf{Z}\tilde{\mathbf{h}}\|^2 &= \tilde{\mathbf{h}}^H(\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H \otimes \mathbf{V}\mathbf{\Gamma}\mathbf{V}^H)\tilde{\mathbf{h}} \\ &= \boldsymbol{\xi}^H(\mathbf{\Lambda} \otimes \mathbf{\Gamma})\boldsymbol{\xi} \end{aligned}$$

where $\boldsymbol{\xi} = (\mathbf{U}^H \otimes \mathbf{V}^H)\tilde{\mathbf{h}}$. Substituting this back into the probability for the deep fade event, and keeping only those products of eigenvalues $\lambda_i\gamma_d$ that are nonzero:

$$P_{deep}(SNR|\boldsymbol{\mathcal{R}}) = \Pr(\sum_{k=1}^{\eta\eta_L} SNR\psi_k|\xi_k|^2 \leq 1|\boldsymbol{\mathcal{R}}) \tag{C.3}$$

Each of the random variables $|\xi_k|^2$ is exponentially distributed. As shown in [70]:

$$P_{deep}(SNR|\boldsymbol{\mathcal{R}}) = \sum_{k=1}^{\eta\eta_L} \mathbf{\Pi}_k(1 - e^{\frac{1}{SNR\psi_k}}) \tag{C.4}$$

where $\mathbf{\Pi}_k = \prod_{l\neq k}^{\eta\eta_L} \psi_k/(\psi_k - \psi_l)$. Taking a Taylor series expansion of the exponential term:

$$\begin{aligned}
P_{deep}(SNR|\boldsymbol{\mathcal{R}}) &= -\sum_{p=1}^{\infty} \frac{1}{p!}\left(-\frac{1}{SNR}\right)^p \sum_{k=1}^{\eta\eta_L} \frac{\psi_k^{\eta\eta_L-1-p}}{\prod_{l\neq k}(\psi_k-\psi_l)} \\
&\overset{(a)}{=} -\sum_{p=\eta\eta_L}^{\infty} \frac{1}{p!}\left(-\frac{1}{SNR}\right)^p \sum_{k=1}^{\eta\eta_L} \frac{\psi_k^{\eta\eta_L-1-p}}{\prod_{l\neq k}(\psi_k-\psi_l)} \\
&\overset{(b)}{=} -\sum_{p=\eta\eta_L}^{\infty} \frac{1}{p!}\left(-\frac{1}{SNR}\right)^p \left(\prod_{r=1}^{\eta\eta_L} \frac{1}{\psi_r^{p-\eta\eta_L+1}}\right) \sum_{k=1}^{\eta\eta_L} \prod_{l\neq k} \frac{\psi_l^{p-\eta\eta_L+1}}{\psi_k-\psi_l} \\
&\overset{(c)}{=} -\sum_{p=\eta\eta_L} \frac{1}{p!}\left(-\frac{1}{SNR}\right)^p \left(\prod_{r=1}^{\eta\eta_L} \frac{1}{\psi_r^{p-\eta\eta_L+1}}\right) \sum_{k=1}^{\eta\eta_L} \prod_{l\neq k} \frac{1}{\frac{\psi_k}{\psi_l^{p-\eta\eta_L+1}} - \frac{1}{\psi_l^{p-\eta\eta_L}}} \\
&\overset{(d)}{=} \frac{1}{p!}\left(\prod_{r=1}^{\eta\eta_L} \frac{1}{SNR\psi_r}\right) \sum_{k=1}^{\eta\eta_L} \prod_{l\neq k} \frac{\psi_l}{\psi_k-\psi_l} + O(SNR^{-(N+1)})
\end{aligned}$$

Step (a) follows by first noting:

$$\sum_{k=1}^{N} \frac{\psi_k^m}{\prod_{l\neq k}(\psi_k-\psi_l)} = \begin{cases} 0 & 0 \leq m < N-1 \\ 1 & m = N-1 \\ \sum_{k=1}^{N} \psi_k & m = L \end{cases} \tag{C.5}$$

thus the first $\eta\eta_L - 1$ elements of the summation ($p = 1, ..., \eta\eta_L - 1$) will be zero. Step (b) is a result of the fact:

$$\begin{aligned}
\sum_{k=1}^{\eta\eta_L} \frac{\psi_{k^{\eta\eta_L-1-p}}}{\prod_{l\neq k}(\psi_k-\psi_l)} &= \sum_{k=1}^{\eta\eta_L} \frac{1}{\psi_k^{p-\eta\eta_L+1}\prod_{l\neq k}(\psi_k-\psi_l)} \\
&= \prod_{r=1}^{\eta\eta_L} \frac{1}{\psi_r^{p-\eta\eta_L+1}} \sum_{k=1} \prod_{k\neq l} \frac{\alpha_l^{p-\eta\eta_L+1}}{\alpha_k-\alpha_l}
\end{aligned} \tag{C.6}$$

Step (c) is a simple division by $\psi_l^{p-\eta\eta_L+1}$ in the last term, while step (d) just keeps the $p = \eta\eta_L$-th element of the summation, noting that the remaining elements are much smaller for high SNR. The final step can be done via Lagrange interpolating polynomials: for a given set of points (y_k, x_k) the unique Lagrange interpolating polynomial for these points is given by:

$$y(x) = \sum_{k=1} y_k \prod_{l\neq k} \frac{x-x_k}{x_k-x_l} \tag{C.7}$$

This is precisely the last term in the expression, where $y_k = (-1)^{\eta\eta_L - 1}$ and $x = 0$. Since the Lagrange interpolating polynomial of two points is a line, and all the points we have specified remain on that line, the lagrange interpolating polynomial of these $\eta\eta_L$ points is a line, and thus when evaluated at $x = 0$, $y(0) = (-1)^{\eta\eta_L - 1}$, implying:

$$\sum_{k=1}^{\eta\eta_L} \prod_{l \neq k} \frac{\psi_l}{\psi_k - \psi_l} = (-1)^{\eta\eta_L - 1} \tag{C.8}$$

Finally, the probability of the deep fade event:

$$P_{deep}(SNR|\boldsymbol{\mathcal{R}}) = \frac{\prod_{k=1}^{d^*(\mathcal{R})} \psi_k^{-1}}{d^*(\boldsymbol{\mathcal{R}})!} SNR^{-d^*(\mathcal{R})} + O(SNR^{-d^*(\mathcal{R})-1}) \tag{C.9}$$

Substitution of (C.9) into (C.1) proves our lower bound.

The upper bound on the probability of error is determined using the union bound on the worst pairwise error probability, obtained as in [86]:

$$\begin{aligned} P_e(SNR|\boldsymbol{\mathcal{R}}) &\leq (|\mathcal{M}| - 1) \max_{i,j} \mathbb{E}_{\tilde{\mathbf{h}}} \left\{ \exp(-\|\mathbf{A}_{ij}\mathbf{Z}\tilde{\mathbf{h}}\|^2 \frac{SNR}{4}) \right\} \\ &= (|\mathcal{M}| - 1) \max_{i,j} \det \left(\mathbf{I} + \frac{SNR}{4} \mathbf{A}_{ij}^H \mathbf{A}_{ij} \mathbf{Z}\mathbf{Z}^H \right)^{-1} \end{aligned} \tag{C.10}$$

We will use the following bound found in [34]. For positive semi-definite matrices $\mathbf{A}^H\mathbf{A}$ and $\mathbf{Z}\mathbf{Z}^H$ with ordered eigenvalues $\alpha_1 \geq ... \geq \alpha_{D(L+1)}$ and $\psi_1 \geq ... \geq \psi_{D(L+1)}$ respectively,

$$\det\left((\mathbf{A}^H\mathbf{A})^{-1} + \mathbf{Z}\mathbf{Z}^H \frac{SNR}{4} \right) \geq \prod_{k=1}^{D(L+1)} \left(\alpha_{D(L+1)-k+1}^{-1} + \frac{SNR}{4} \psi_k \right) \tag{C.11}$$

Assuming $\mathbf{A}_{ij}^H\mathbf{A}_{ij}$ is full rank, and SNR is sufficiently large

$$\begin{aligned} P_e(SNR|\boldsymbol{\mathcal{R}}) &\leq (|\mathcal{M}| - 1) \max_{i,j} \prod_{k=1}^{d^*(\mathcal{R})} \left(1 + \frac{SNR}{4} \psi_k \alpha_{D(L+1)-k+1}^{ij} \right)^{-1} \\ &\leq (|\mathcal{M}| - 1) \prod_{k=1}^{d^*(\mathcal{R})} \psi_k^{-1} \left(\frac{SNR}{4} \right)^{-d^*(\mathcal{R})} \\ &\quad \cdot \max_{i,j} \prod_{l=1}^{d^*(\mathcal{R})} (\alpha_{D(L+1)-l+1}^{ij})^{-1} \end{aligned} \tag{C.12}$$

C.2 Proof of Theorem 4.3.3

Let us assume that $D \leq T$. Then the probability density function of $\mathbf{R}\mathbf{R}^H$ is

$$f_{\mathbf{R}\mathbf{R}^H}(\mathbf{Z}) = \frac{1}{c_{D,T}} \int_{\Gamma} p(\sqrt{\mathbf{Z}}\tilde{\mathbf{U}}) \det(\mathbf{Z})^{T-D} \mu(d\mathbf{U})$$

where the $D \times D$ matrix $\mathbf{Z} \in \Lambda$, $\mathbf{U} = (u_{ik}) \in \Gamma$, $\tilde{\mathbf{U}} = (u_{ik})_{i=1\ldots D,\ k=1\ldots T}$, $c_{D,T} = \pi^{D(D-1)/2-DT} \prod_{k=1}^{D} (T-k)!$ [40]. Also, Γ denotes the set of unitary $T \times T$ matrices and μ is the normalized Haar measure on it, and Λ denotes the set of Hermitian positive definite matrices. By using the formula for the pdf of $\mathbf{R}\mathbf{R}^H$, we conclude that $E\{\det(\mathbf{R}\mathbf{R}^H)^{-1}\} < \infty$ if and only if the integral

$$I \triangleq \int_{\Lambda} \int_{\Gamma} p(\sqrt{\mathbf{Z}}\tilde{\mathbf{U}}) \det(\mathbf{Z})^{T-D-1} \mu(d\mathbf{U}) d\mathbf{Z}$$

is finite. The notation $d\mathbf{Z}$ refers to the Lebesgue measure on the set of $D \times D$ dimensional matrices. The proof follows by bounding the integral I under the given conditions of the Theorem. See [82, Theorem 3] for a similar proof for the flat fading case, i.e., $\eta_L = 1$. For $D \geq T$, the proof is extended by replacing $\mathbf{R}\mathbf{R}^H$ by $\mathbf{R}^H\mathbf{R}$.

BIBLIOGRAPHY

[1] Use of High Frequency Ionospheric Channel Simulators. 1978. International Telecommunications Union, Radio Communications Sector (ITU-R) Recommendation F.520-2.

[2] Interoperability and Performance Standards for Data Modems. 2000. Department of Defense MIL-STD-188-110B.

[3] Testing of HF Modems with Bandwidths of up to 12kHz Using Ionospheric Channel Simulator. 2000. International Telecommunications Union, Radio Communications Sector (ITU-R) Recommendation F.1487 .

[4] S. Adireddy, L. Tong, and H. Viswanathan. Optimal placement of training for frequency-selective block-fading channels. *IEEE Trans. Info. Theory*, 48(8):2338–2353, August 2002.

[5] A. M. Alamouti. A Simple Transmit Diversity Technique for Wireless Communications. *IEEE Trans. Commun.*, 16:1451–1458, October 1998.

[6] C. D. Austin, R. L Moses, J. N. Ash, and E. Ertin. On the relation between sparse reconstruction and parameter estimation with model order selection. *IEEE Journal of Selected Topics in Signal Processing*, 4(3):560–570, 2010.

[7] W. U. Bajwa, J. Haupt, G. Raz, and R. Nowak. Compressed channel sensing. In *Proc. CISS*, March 2008.

[8] W. U. Bajwa, A. M. Sayeed, and R. Nowak. Compressed sensing of wireless channels in time, frequency, and space. In *Proc. of Asilomar Conf. Signals, Systems, and Computers*, pages 1–5, Oct 2008.

[9] W. U. Bajwa, A. M. Sayeed, and R. Nowak. Learning sparse doubly-selective channels. In *Allerton '08*, September 2008.

[10] I. Barhun, G. Leus, and M. Moonen. MMSE estimation of basis expansion models for rapidly time-varying channels. In *Proc. EUSIPCO '05*, 2005.

[11] P. A. Bello. Characterization of randomly time-variant linear channels. *IEEE Trans. Comm.*, 11(4):360–393, December 1963.

[12] C. R. Berger, S. Zhou, W. Chen, and P. Willett. Sparse channel estimation for ofdm: Over-complete dictionaries and super-resolution. In *Signal Processing Advances in Wireless Communications*, pages 196–200, 2009.

[13] C. R. Berger, S. Zhou, J. Preisig, and P. Willett. Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing. *IEEE Transactions on Signal Processing*, 58(3):1708–1721, March 2010.

[14] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman. A simple cooperative diversity method based on network path selection. *IEEE Journal on Selected Areas in Communications*, 24(3):659–672, March 2006.

[15] T. Blu, P. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot. Sparse sampling of signal innovations. *IEEE Sig. Proc. Magazine*, 25(2):31–40, March 2008.

[16] D. K. Borah and B. D. Hart. Frequency-selective fading channel estimation with a polynomial time-varying channel model. *IEEE Trans. Comm.*, 47:862–873, June 1999.

[17] A. M. Bruckstein, D. L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*, 51(1):34–81, February 2009.

[18] E. Candes, Y. Eldar, and D. Needell. Compressed sensing with coherent and redundant dictionaries. Preprint, May 2010.

[19] E. J. Candes, J. Romberg, and T. Tao. Stable signal recovery from incomplete and innacurate measurements. *Comm. Pure Appl. Math*, 59(8):1207–1223, 2006.

[20] E. J. Candes and T. Tao. Decoding by linear programming. *IEEE Trans. Info. Theory*, 51(12):4203–4215, 2005.

[21] E. J. Candes and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Info. Theory*, 53(12):5406–5425, 2006.

[22] E. J. Candes and T. Tao. The Dantzig Selector: Statistical estimation when p is much larger than n. *Ann. Stat.*, 35:2313–2351, 2007.

[23] J. K. Cavers. An analysis of pilot symbol assisted modulation for Rayleigh fading channels. *IEEE Trans. Veh. Technol.*, 40(4):686–693, November 1991.

[24] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scient. Comput.*, 20:33–61, 1998.

[25] A. P. Clark and F. McVerry. Channel estimation for an HF radio link. *IEE Proceedings in Communications Radar and Signal Processing*, 128(1):33–42, February 1981.

[26] S. F. Cotter and B. D. Rao. Sparse channel estimation via matching pursuit with application to equalization. *IEEE Trans. Comm.*, 50(3):374–377, March 2002.

[27] M. Dong and L. Tong. Optimal design and placement of pilot symbols for channel estimation. *IEEE Trans. on Sig. Proc.*, 50(12):3055–3069, December 2002.

[28] D. Donoho and M. Elad. Optimally sparse representations in arbitrary redundant bases. In *Proc. National Acad. Sci. USA*, volume 100, pages 2197–2202, March 2003.

[29] D. Donoho, M. Elad, and V. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Info. Theory*, 52(1):6–18, January 2006.

[30] D. Donoho and J. Tanner. Sparse nonnegative solution of underdetermined linear equations by linear programming. In *Proc. National Academy of Sciences*, volume 102, pages 9446–9451, 2005.

[31] P. J. Durka, A. Matysiak, and E. Martinez Montes. Multichannel matching pursuit and EEG inverse solutions. *Journal of Neuroscience Methods*, 148(1):44–59, 2005.

[32] D. Eiwen, G. Taubock, F. Hlawatsch, and H. G. Fiechtinger. Group sparsity methods for compressive channel estimation in doubly-dispersive multicarrier systems. In *IEEE SPAWC 2010*, June 2010.

[33] D. Eiwen, G. Taubock, F. Hlawatsch, H. Rauhut, and N. Czink. Multichannel-compressive estimation of doubly-selective channels in MIMO-OFDM systems: Exploiting and enhancing joint sparsity. In *Proc. IEEE ICASSP*, pages 3082–3085, March 2010.

[34] M. Fiedler. Bounds for the determinant of the sum of hermitian matrices. In *Proc. of the American Mathematicl Society*, volume 30, pages 27–31, September 1971.

[35] G. J. Foschini and M. J. Gans. On limits of wireless communications in a fading environment when using mulitple antennas. *Wireless Personal Communications*, 6:311–335, 1998.

[36] J. J. Fuchs. Multipath time-delay estimation. In *Proc. IEEE ICASSP*, pages 527–530, 1997.

[37] J. J. Fuchs. On sparse representations in arbitrary redundant bases. *IEEE Trans. Info. Theory*, 50(6):1341–1344, 2004.

[38] S. Geirhofer, L. Tong, and A. Scaglione. Time-reversal space-time coding for doubly-selective channels. In *IEEE WIreless Communications and Networking Conf. (WCNC)*, volume 3, pages 1638–1643, April 2006.

[39] G. B. Giannakis and C. Tepedelenlioglu. Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels. In *Proc. IEEE*, pages 1969–1986, October 1998.

[40] V. Girko. *Theory of Random Determinants*. Kluwer Academic Publishers, 1990.

[41] I. F. Gorodnitsky and B. D. Rao. Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm. *IEEE Trans. Sig. Proc.*, 45(3):600–616, March 1997.

[42] R. Gribonval. Piecewise linear source separation. In *Proc. SPIE*, volume 5207, San Diego, CA, 2003.

[43] S. Hariharan and A. P. Clark. Hf channel estimation using a fast transversal filter algorithm. *IEEE Trans. on Acoustics Speech and Signal Processing*, 38(8):1353–1362, August 1990.

[44] S. He and J. K. Tugnait. Doubly-selective channel estimation using exponential basis models and subblock tracking. In *Proc. GLOBECOM '07*, pages 2847–2851, November 2007.

[45] M. A. Herman and T. Strohmer. High-resolution radar via compressed sensing. *IEEE Trans. Sig. Proc.*, 57(6):2275–2284, 2009.

[46] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[47] F. M. Hsu. Square root kalman filtering for high-speed data received over fading dispersive HF channels. *IEEE Trans. Info. Theory*, 28(5):753–763, September 1982.

[48] R. A. Iltis. Joint estimation of PN code delay and multipath using extended Kalman filter. *IEEE Trans. Comm.*, 38(10):1677–1685, Oct 1990.

[49] A. P. Kannu and P. Schniter. Design and analysis of MMSE pilot-aided cyclic-prefixed block transmission for doubly-selective channels. *IEEE Trans. Sig. Proc.*, 56(3):1148–1160, March 2008.

[50] M. Kocic, D. Brady, and M. Stojanovic. Sparse equalization for real-time digital underwater acoustic communications. In *IEEE OCEANS '95*, volume 3, pages 1417–1422, October 1995.

[51] H. Krim and M. Viberg. Two decades of array signal processing research. *IEEE Sig. Proc. Magazine*, pages 67–94, July 1996.

[52] J. B. Kruskal. Three-way arrays: Rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics. *Linear Algebra and its Applications*, 18(2):95–138, 1977.

[53] J. N. Laneman, D. N. C. Tse, and G. W. Wornell. Cooperative diversity in wireless networks: Efficient protocols and outage behavior. *IEEE Trans. Info. Theory*, 50(12):3062–3080, December 2004.

[54] J. N. Laneman and G. W. Wornell. Distributed space-time coded protocols for exploiting cooperative diversity in wireless networks. *IEEE Trans. Info. Theory*, 49(10):2415–2425, Oct. 2003.

[55] E. G. Larsson and P. Stoica. *Space-Time Block Coding For Wireless Communications*. Cambridge University Press, 2003.

[56] K. F. Lee and D. B. Williams. A Space-Frequency Transmitter Diversity Technique for OFDM Systems. In *GLOBECOM*, volume 3, pages 1473–1477, December 2000.

[57] W. Li and J. Preisig. Estimation of rapidly time-varying sparse channels. *IEEE Journal of Oceanic Engineering*, 32(4):927–939, Oct 2007.

[58] X. Li. Space-time coded multi-transmission among distributed transmitters without perfect synchronization. *IEEE Signal Processing Letters*, 11(12):948–951, December 2004.

[59] E. Lindskog and A. Paulraj. A transmit diversity scheme for channels with intersymbol interference. In *IEEE ICC*, volume 1, pages 307–311, June 2000.

[60] T. Liu and D. K. Borah. Estimation of time-varying frequency-selective channels using a matching pursuit technique. In *IEEE WCNC*, pages 941–946, 2003.

[61] L. Ljung. *System Identification: Theory for the User*. Prentice Hall, 2 edition, 1999.

[62] X. Ma, G. B. Giannakis, and S. Ohno. Optimal training for block transmissions over doubly selective wireless fading channels. *IEEE Trans. Sig. Proc.*, 51(5):1351–1366, May 2003.

[63] D. Malioutov, M. Cetin, and A. Willsky. A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Trans. Sig. Proc.*, 53(8):3010–3022, August 2005.

[64] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Sig. Proc.*, 41:3397–3415, December 1993.

[65] A. F. Molisch. Ultrawideband propagation channels- theory, measurement, and modeling. *IEEE Trans. Veh. Technol.*, 54(5):1528–1545, September 2005.

[66] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayargue. Subspace methods for the blind identification of multichannel FIR filters. *IEEE Trans. on Sig. Proc.*, 43:516–525, Feb. 1995.

[67] R. Negi and J. Cioffi. Pilot tone selection for channel estimation in a mobile OFDM system. *IEEE Trans. Veh. Tech.*, 44(3):1122–1128, 1998.

[68] S. Ohno and G. B. Giannakis. Optimal training and redundant precoding for block tranmissions with application to wireless OFDM. *IEEE Trans. Comm.*, 50(12):2113–2123, December 2002.

[69] Y. C. Pati, R. Rezaiifar, and P.S. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decom-

position. In *Asilomar Conf. on Signals, Systems, and Computers*, volume 1, pages 40–44, 1993.

[70] J. G. Proakis and M. Salehi. *Digital Communications*. McGraw Hill, 5th edition, 2008.

[71] Mark A. Richards. *Fundamentals of Radar Signal Processing*. McGraw Hill, 2005.

[72] M. Rudelson and R. Vershynin. Sparse reconstruction by convex relaxation: Fourier and gaussian measurements. In *Proc. CISS*, pages 207–212, March 2006.

[73] J. J. Rushanan. On the vandermonde matrix. *The American Mathematical Monthly*, 96(10):921–924, 1989.

[74] A. Scaglione and Y. W. Hong. Oppurtunistic large arrays: Cooperative transmission in wireless multihop ad-hoc networks to reach far distances. *IEEE Trans. Sig. Proc.*, 51(8):2082–2092, August 2003.

[75] P. Schniter, L. C. Potter, and J. Ziniel. Fast bayesian matching pursuit: Model uncertainty and parameter estimation for sparse linear models. In *Info. Theory and App. Workshop*, pages 326–333, 2008.

[76] W. F. Schreiber. Advanced television systems for terrestrial broadcasting: Some problems and some proposed solutions. In *Proc. of IEEE*, volume 83, pages 958–981, June 1995.

[77] A. Sendonaris, E. Erkip, and B. Aazhang. User cooperation diversity - part 1: System description; part 2: Implementation aspects and performance. *IEEE Trans. Comm.*, 51(11):1927–1948, 2003.

[78] M. Sharp and A. Scaglione. Distributed randomized space-time coding for HF transmission. In *MILCOM '06*, pages 1–6, October 2006.

[79] M. Sharp and A. Scaglione. Application of sparse signal recovery to pilot-assisted channel estimation. In *Proc. IEEE ICASSP*, pages 3469–3472, March 2008.

[80] M. Sharp and A. Scaglione. Estimation of sparse multipath channels. In *MILCOM '08*, November 2008.

[81] M. Sharp and A. Scaglione. On time varying channel estimation using sparse models. In *IEEE ICC '10*, Cape Town, South Africa, May 2010.

[82] B. Sirkeci-Mergen and A. Scaglione. Randomized space-time coding for distributed cooperative communication. *IEEE Trans. on Sig. Proc.*, 55(10):5003–5017, Oct. 2007.

[83] P. Stevenhagen and H. Lenstra. Chebotarev and his density theorem. *Math. Intell.*, 18(2):26–37, 1996.

[84] T. Tao. An uncertainty principle for cyclic groups of prime order. *Math. Res. Letter*, 12:121–127, 2005.

[85] V. Tarokh, H. Jafarkhani, and A. Calderbank. Space-time block codes from orthogonal designs. *IEEE Trans. Info. Theory*, 45(5):1456–1467, July 1999.

[86] V. Tarokh, N. Seshadri, and A. R. Calderbank. Space-time codes for high data rate wireless communication: Performance criterion and code construction. *IEEE Trans. Info. Theory*, 44(2):744–765, March 1998.

[87] G. Taubock and F. Hlawatsch. A compressed sensing technique for OFDM channel estimation in mobile environments: Exploiting channel sparsity for reducing pilots. In *Proc. IEEE ICASSP*, 2008.

[88] G. Taubock, F. Hlawatsch, D. Eiwen, and H. Rauhut. Compressive estimation of doubly-selective channels in multicarrier systems: Leakage effects and sparsity-enhancing processing. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):255–271, April 2010.

[89] R. Tibshirani. Regression shrinkage and selection via LASSO. *J. R. Statist. Soc. B*, 58(1):267–288, 1996.

[90] M. E. Tipping. Sparse bayesian learning and the relevance vector machine. *Journal Machine Learning Res.*, 1:211–244, 2001.

[91] J. A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Info. Theory*, 50(10):2231–2242, Oct. 2004.

[92] M. K. Tsatsanis, G. B. Giannakis, and G. Zhou. Estimation and equalization of fading channels with random coefficients. In *IEEE ICASSP '96*, volume 2, pages 1093–1096, May 1996.

[93] D. Tse and P. Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, Cambridge, U.K., 2005.

[94] M. Vetterli, P. Marziliano, and T. Blu. Sampling signals with finite rate of innovation. *IEEE Trans. Sig. Proc.*, 50:1417–1428, 2002.

[95] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath. Optimal training for frequency-selective fading channels. In *Proc. IEEE ICASSP*, volume 4, pages 2105–2108, May 2001.

[96] M. Visintin. Karhunen-Loeve expansion of a fast Rayleigh fading process. *IEEE Electronics Letters*, 32:1712–1713, August 1996.

[97] C. C. Watterson, J. R. Jurosheck, and W. D. Bensema. Experimental confirmation of an HF channel model. *IEEE Transactions on Communication Technology*, 18:792–803, December 1970.

[98] S. Wei, D. Goeckel, and M. Valenti. Asynchronous cooperative diversity. *IEEE Trans. on Wireless Communications*, 5(6):1547–1557, June 2006.

[99] S. Yiu, R. Schober, and L. Lampe. Distributed space-time block coding. *IEEE Trans. Comm.*, 54(7):1195–1206, July 2006.